Spatial stochastic point processes.

Lotwick, H. W.

Award date:
1981

Awarding institution:
University of Bath

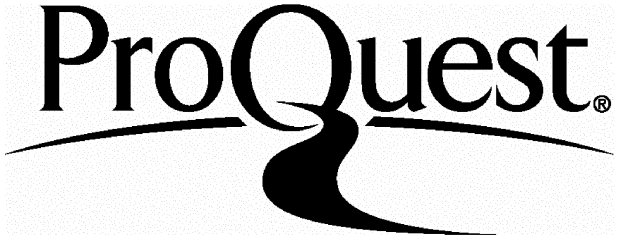
SPATIAL STOCHASTIC POINT PROCESSES

submitted by H.W. Lotwick for the
degree of Ph.D. of the University of Bath

1981

H.W. Lotwick

H.W. Lotwick



ABSTRACT

This thesis is concerned with some practical and theoretical problems in the field of spatial stochastic point processes. The work described falls into two main areas: Chapters 1-3 describe some techniques for analysing spatial patterns, while, in Chapters 4-7, properties of certain stochastic models for spatial patterns are investigated, theoretically and empirically. We consider simple point processes in the plane, strictly stationary under rigid motions of the plane; many of the techniques and results can be generalised to more than two dimensions.

Chapter 1 discusses the use of empty space in analysing spatial patterns; the method described estimates the probability that a randomly placed set of a given shape and size is empty of points of the pattern. Chapter 2 describes the use of histograms to estimate $g(t)/\lambda^2$ for a spatial pattern; this technique is of use during the earlier stages of the analysis of a pattern. In Chapter 3, empty space techniques are generalised for the analysis of 'multitype' point patterns, and are compared with second-order methods for analysing such patterns. All the techniques of Chapters 1-3 are illustrated by examples.

Chapters 4-7 deal mainly with models for 'hard core' point patterns. Chapter 4 contains a review of various models for hard core processes. Chapters 5 and 6 are concerned with spatial birth and death processes which are used to simulate realisations of certain models for point patterns, notably Kelly-Ripley models. In Chapter 5, coupling techniques from Markov Chain theory are used to obtain theoretical convergence results for such processes. Chapter 6 describes a computer algorithm for simulating hard core birth and death processes at high packing densities. In

Chapter 7, this algorithm is modified to simulate the 'SSI' process and to investigate the complete packing of non-overlapping discs in a rectangular container.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr Bernard Silverman, for his valuable guidance and encouragement throughout the course of this work. I would also like to thank Professor Robin Sibson and Dr Tim Brown for many valuable discussions, and for the provision of computer algorithms.

The research described here was supported by a Social Science Research Council Research Studentship and a Bath University Research Studentship.

Finally, I an indebted to Ruth Nessbert for her accurate and speedy typing.

NOTATION

The following notation is used in parts of this thesis.

$Z_R(x)$	The closed disc of radius R centred at the point x.
$d(x,y)$	The Euclidean distance in the plane between the points x and y. (This is occasionally denoted by $\|x - y\|$).
$m(A)$	The (two-dimensional) Lebesgue measure of the set A. (This is occasionally denoted by $\|A\|$).
$\|M\|$	The number of elements in the finite set M.
∂A	The boundary of the Borel set A.
$N_x(A)$	The number of points from the pattern of x-points that lie in the set A.
$A \searrow r$	The set $\{x \in A: Z_r(x) \subseteq A\}$, where A is a Borel set, and $r \in \mathbb{R}$.

CONTENTS

CHAPTER 0

PRELIMINARY DISCUSSION

1. Introduction

This thesis is concerned with some practical and theoretical problems in the field of spatial stochastic point processes. The work described falls into two main areas. Chapters 1, 2 and 3 are concerned with some techniques for analysing spatial patterns, while, in Chapters 4-7, properties of certain stochastic models for spatial point patterns are investigated, both theoretically and empirically.

All the point processes considered lie in the plane, though many of the techniques described and results obtained can be generalised to more than two dimensions. We consider simple point processes strictly stationary under rigid motions of the plane (see Ripley, 1976, for details).

2. Methods of Analysis for Spatial Point Patterns

The intensity, λ, of a point process is the expected number of points per unit area; this parameter summarises the first order structure of the process. Ripley (1976 and 1977) introduced the function K(t) to describe the second order structure of a point process; this can be described intuitively as follows.

i) $\lambda^2 K(t)$ is the expected number of ordered pairs of distinct points less than distance t apart with the first point in a given set of unit area.

ii) $\lambda K(t)$ is the expected number of further points within t of an arbitrary point of the process.

iii) Under additional assumptions, $g(t) = \frac{\lambda^2}{2\pi t} \frac{dK}{dt}$ is a joint density for the occurrence of two points distance t apart.

Other properties of a point process can be summarised using Ripley's function p(t) (Ripley 1977); p(t) is defined to be the probability that a disc of radius t contains at least one point of the process. Both K and p are, of course, 'cumulative functions'.

Suppose that we observe a point pattern $x = \{x_1, \ldots, x_N\}$ in a sampling window E in the plane, being that part of a realisation of a point process X lying inside E. Ripley (1976 and 1977) defined estimators $\hat{K}$ and $\hat{p}$ of the functions K and p. These can be described as follows. For $t < t_o = \inf\{ s: \exists\, x \in E,\ \partial Z_x(s) \cap E = \emptyset \}$, define

$$\hat{K}(t) = \frac{m(E)}{N(N-1)} \sum_i \sum_{j \neq i} k(x_i, x_j)\, I(d(x_i, x_j) < t). \qquad (0.1)$$

Here, k(x,y) is the reciprocal of the proportion of the perimeter of the circle centred on x and passing through y which is within the sampling

window. $\hat{K}$ is thus a weighted empirical distribution function of the inter-point distances of the pattern. Ripley originally used N^2 rather than N(N-1) in the denominator of $\hat{K}$; use of N(N-1) is more natural, and has the advantage that, when X is a Poisson process, $\hat{K}(t)$ is, conditional on the number, N, of points lying in E, an unbiased estimator of K(t).

Ripley's estimator $\hat{p}(t)$ can be described as follows. For a finite subset α of E,

$$\hat{p}(t) = |\{x \in \alpha: (x + tB) \cap (\text{data}) \neq \emptyset\}|/|\alpha| \qquad (0.2)$$

Here, $B = Z_1(\underset{\sim}{0})$, the unit disc centred at the origin, and α is chosen so that $x + tB \subseteq E \ \forall \ x \in \alpha$. α is usually taken as either a regular grid of points or a random sample of points from $\{x: x + tB \subseteq E\}$. $\hat{p}(t)$ is easily seen to be an unbiased estimator of p(t).

When analysing a spatial pattern, $\hat{K}(t)$ (or $\sqrt{\hat{K}(t)}$) and $\hat{p}(t)$ are usually plotted against t. Hypothesis testing and model fitting using $\hat{K}$ or $\hat{p}$ are generally based on Monte Carlo simulation - see Ripley (1977) for examples. Since its introduction, much use has been made of $\hat{K}$ in the analysis of spatial patterns.

3. Some Models for Spatial Point Patterns

In this section, for reference purposes, models for spatial patterns, used elsewhere in the thesis, are described, and some of their properties are given. The list of models given is by no means complete, but it describes all the models used in the remainder of the thesis. Ripley (1977, section 3) gives another list of models for spatial point patterns. Some of the processes described below can be defined on the whole of the plane, while others can only be defined on a bounded region.

i) The Poisson Process

This is the basic model, providing the simplest stochastic model for a spatial point pattern. It is defined by the following properties: if, for a bounded Borel set A, we denote by N(A) the number of points in A, then N(A) has a Poisson distribution, and $\{N(A_i)\}$ are independent whenever $\{A_i\}$ is a disjoint class of such sets. It follows that, if λ is the intensity of the process, then N(A) has mean $\lambda m(A)$, where m(.) denotes Lebesgue measure, and $K(t) = \pi t^2$ irrespective of λ. Conditional on $\{N(A) = n\}$, the n points in A will be independently and uniformly distributed within A.

The Poisson process is a model for 'spatial randomness', and the null hypothesis of an underlying Poisson process is often tested for a spatial pattern. As well as acting as a foundation on which more complex models may subsequently be built, the Poisson process, with its property that there is no interaction between its points, provides a reference with which spatial patterns can be compared. Patterns in which there is attraction between points are said to be

'clustered', whereas ones in which the points repel each other are called 'regular' of 'self-inhibiting'. The null hypothesis of an underlying Poisson process can thus be viewed as a 'dividing' hypothesis in the sense of Cox (1977).

ii) Clustered processes

We consider Neyman-Scott cluster processes; these are based on the Poisson process. Randomly distributed parents are assumed to generate random numbers of offspring which are spatially distributed about the corresponding parent. Formally, we make three assumptions:

Parent points form a Poisson process of intensity μ.

Each parent produces, independently, a random number M of 'offspring' points.

Each offspring point is, independently, spatially distributed relative to the corresponding parent.

We usually assume that parents and offspring are indistinguishable, so that the final pattern consists of both parents and offspring; occasionally, we assume that the parents are not visible, in which case the pattern consists of the offspring only.

All the cluster processes considered in this thesis will be of a particular kind, where the offspring lie within a disc of fixed radius centred at the corresponding parent; we shall call these Poisson cluster processes. We shall often be concerned with a special class of Poisson cluster processes, which we shall call Poisson-Poisson cluster processes. Under this model, the number M of offspring produced by each parent is a Poisson random variable with mean λ; these offspring are then independently and uniformly distributed in a disc of radius R centred at the corresponding parent. A Poisson-Poisson cluster process

thus has three parameters, (μ, λ, R).

Another commonly used Neyman-Scott cluster process is the Gaussian cluster process, in which the number of offspring per parent is Poisson, and the spatial distribution of each offspring about its parent is a symmetric radial normal with probability density function

$$h(\underset{\sim}{x}) = (2\pi\sigma^2)^{-1} \exp(-\tfrac{1}{2}x^T x/\sigma^2) \qquad (|x| < \infty)$$

(see Diggle, 1979).

Neyman-Scott models are reasonably tractable; expressions can be obtained for g(t) and K(t) for these models - see Ripley (1977).

Methods for simulating cluster processes are described by Ripley (1979b) and Diggle (1979); in each case, the number, I, of clusters, and the total number, N, of points in the sampling window D are fixed for the purposes of the simulation. Ripley terms his simulated processes Matérn cluster processes. Note that Diggle uses periodic boundary conditions for his simulations.

iii) Self-inhibiting processes

a) Hard core processes

These are extreme cases of self-inhibiting processes; the points may be regarded as the centres of non-overlapping discs of fixed diameter R, so that no two points may lie within distance R of each other. Various models for hard core processes are discussed in detail in Chapter 4.

b) Kelly-Ripley processes

These models were introduced by Strauss (1975) and corrected by Kelly and Ripley (1976). The points lie in a bounded region, and we

define the process by the Radon-Nikodym derivative of its distribution with respect to that of a Poisson process; henceforth, we shall call this the 'joint density' of a pattern. For the Kelly-Ripley model, the joint density of a pattern x is proportional to

$$f(x) = b^n c^s$$

where n = number of points in the configuration **x**, and s = number of pairs of points within distance R of each other; here, R, b and c are constants with $R > 0$, $b > 0$ and $0 \leqslant c \leqslant 1$. The case $c = 0$ is one model for a hard core process. Kelly-Ripley processes are usually simulated using spatial birth and death processes, as described by Ripley (1977 and 1979a). Some aspects of these birth and death processes are discussed in Chapters 5 and 6 of this thesis.

Kelly-Ripley models are a special case of pairwise interaction models in which the interaction between pairs of points depends on the distance between the pair. Again, the points lie in a bounded region. For such a model, the joint density is proportional to

$$f(x) = b^n \prod_{\substack{\xi,\eta \in x \\ \xi \neq \eta}} h(d(\xi,\eta))$$

where $b > 0$, n is as above, and h is an interaction function defined on $(0,\infty)$ taking values in $[0,\infty)$. Such models are often used for bounded functions h which vanish on $(0,R)$; these conditions ensure that f is integrable.

Pairwise interaction processes are usually simulated using birth and death processes similar to those used for Kelly-Ripley models; some of the results of Chapter 5 apply to such birth and death processes.

4. Outline of the Remaining Chapters of the Thesis

In the first part of this thesis, the use of some techniques other than $\hat{K}$ for spatial analysis is considered. In Chapter One, the use of empty space in the analysis of spatial patterns is discussed. This method, based on the empty spaces in the pattern, estimates the probability that a set of a given shape and size, put down at random, is empty of points of the pattern; it is closely related to Ripley's estimation of p(t). Algorithms for calculating empty space statistics are described, and some asymptotic results for the statistics are proved. The technique is illustrated by examples.

Chapter Two is concerned with the estimation of $g(t)/\lambda^2$, using histograms. This technique would seem to be most useful during the earlier stages of the analysis of a spatial pattern. Such histograms can also be used to estimate the parameters of certain kinds of cluster processes.

Chapter Three is concerned with a different problem - that of analysing a pattern where the points are of several different types. The empty space methods of Chapter One are generalised for the analysis of such multitype patterns. These methods are seen as complementary to second order techniques for multitype patterns. The use of both empty space and second order methods is illustrated on three datasets.

Chapters 4-7 deal mainly with models for 'hard core' point patterns. Chapter Four contains a review of various models for hard core processes. Some of the models are reformulated in terms of random measures, and some properties of the Matérn hard core models are examined.

Chapters Five and Six are concerned with some aspects of spatial birth and death processes which are used to simulate realisations of certain models for point patterns, notably Kelly-Ripley models. In Chapter Five, coupling techniques from Markov Chain theory are used to obtain some theoretical results concerning the convergence to equilibrium of the birth and death processes. In the case of hard core processes, a sufficient condition on the packing density to ensure convergence is established; a lower bound on the rate of convergence to equilibrium is also given.

Chapter Six describes a computer algorithm, based on the Dirichlet Tessellation of the points, for simulating hard core birth and death processes. At very high packing densities, this algorithm is considerably faster than existing methods of simulating the birth and death processes, making it possible to make an empirical study of the convergence properties of the processes at such densities.

In Chapter Seven, the algorithm of Chapter Six is modified in order to simulate the 'SSI' hard core process. The modified algorithm is used to investigate the complete packing of non-overlapping discs in a rectangular container. An estimate is obtained of the limiting value of the expected packing density at complete packing, as the container size tends to infinity.

Some of the material of Chapter Six has appeared in published form, in the following paper

Lotwick, H.W. and Silverman, B.W. (1981): Convergence of Spatial Birth and Death Processes.
Math. Proc. Camb. Phil. Soc., 90, 155-165.

CHAPTER ONE

THE USE OF EMPTY SPACE FOR ANALYSING SPATIAL POINT PATTERNS

1. Introduction

The most commonly used methods in the analysis of spatial data in the form of a point pattern are methods based on the inter-point distances of the pattern. For example, Ripley's $\hat{K}$ (see Ripley, 1977) is much used; Ripley (1979b) gives a survey of distance methods used in spatial analysis. In this chapter, an alternative approach, based on the empty spaces in the point pattern, is investigated.

When a clustered point pattern is inspected visually, it is often the empty spaces or gaps in the point pattern that are most noticeable. The approach described here utilises these empty spaces by estimating the probability that a set of a given shape and size, put down at random, is empty of points of the pattern. Two types of estimate are introduced, and methods of computing them are described. Some asymptotic distributional results are also proved for the empty space estimates; these make it possible to examine various hypotheses about the data. Examples of the use of empty space statistics are given, and, in a simulation study, the effectiveness of empty space methods for the detection of clustering is examined.

2. Empty space probabilities

Let X be a random point pattern in the plane. If A is a measurable subset of $\mathbb{R}^2$, we denote by $N_X(A)$ the number of X-points in A. Define the empty space probability $G_X(A)$ by

$$G_X(A) = P(N_X(A) = 0) \tag{1.1}$$

Thus $G_X(A)$ is the probability that the set A contains no X-points, i.e. is empty of points of the pattern. If (as we will assume throughout) X is stationary under rotations and translations of the plane, $G_X(A)$ depends only on the size and shape of A, and not on its position or orientation. Thus, for example, if we denote by S(t) a square region of side t, and by U(t) a disc of radius t, we can talk of $G_X(S(t))$ and $G_X(U(t))$. In fact, in the case of discs, the empty space probability is equivalent to a property of point processes already described by Ripley (1977), who defined the quantity p(t) to be the probability that a disc of radius t contains at least one point; clearly

$$G_X(U(t)) = 1 - p(t) \tag{1.2}$$

For certain models for random point patterns, it is possible to obtain expressions for $G_X(A)$.

i) If X is a homogeneous Poisson process with parameter λ, then

$$G_X(A) = \exp\{-\lambda m(A)\} \tag{1.3}$$

where m is Lebesgue measure in the plane.

ii) If X is a Poisson-Poisson cluster process, with parameters (μ, λ, R) (see Chapter 0), then

$$G_X(A) = \exp\{-\mu m(A)\} \exp\{-\mu m(A^R \ominus A)\,[1 - f(A^R \ominus A, A)]\}, \tag{1.4}$$

where $A^R = \{x \in \mathbb{R}^2 : d(x,y) \leq R \text{ for some } y \in A\}$

and $$f(A^R \ominus A, A) = \frac{1}{m(A^R \ominus A)} \int_{x \in A^R \ominus A} \exp\{\frac{-\lambda}{\pi R^2} m(A \cap Z_x(R))\} dx \ ;$$

here $Z_x(R) = \{y \in \mathbb{R}^2 : d(x,y) \leq R\}$, and d denotes Euclidean distance in the plane. If the parent process is not visible (see Chapter 0), then (1.4) becomes

$$G_x(a) = \exp\{-\mu m(A^R)[1 - f(A^R, A)]\} \ . \qquad (1.5)$$

iii) For Matérn hard core processes (described in Chapter 4, section 2) expressions can be obtained for $G_x(A)$; these are intractable unless the diameter of the set A is less than the diameter of the discs. When this is the case, it is clear that for any hard core process,

$$G_x(A) = 1 - \lambda \, m \, (A) \ ,$$

where λ is the intensity of the process.

White (1979) notes that the set function $G_x(A)$ depends on the correlation functions of all orders. A statistic based on empty space would thus be an appealing alternative to second order methods for analysing point patterns. In the remainder of this chapter, estimates of empty space probabilities $G_x(A)$ are developed and used for hypothesis testing and model fitting for spatial point patterns.

3. Empty space estimates

Suppose we have a realisation x of a point pattern, lying in a sampling window D in the plane. Let A be a set of a given size and shape, containing the origin; we wish to estimate $G_x(A)$.
For $\underset{\sim}{u}$ a vector in $\mathbb{R}^2$, define

$$A_{\underset{\sim}{u}} = \{\underset{\sim}{z} + \underset{\sim}{u} : \underset{\sim}{z} \in A\} .$$

The statistics used in this chapter to estimate $G_x(A)$ are of two forms. The first of these is

$$\hat{G}_x(A) = \frac{1}{|\alpha|} \sum_{\underset{\sim}{u} \in \alpha} I(N_x(A_{\underset{\sim}{u}}) = 0) \qquad (1.6)$$

where α is a finite subset of $\mathbb{R}^2$, chosen so that $A_{\underset{\sim}{u}} \subseteq D \ \forall \underset{\sim}{u} \in \alpha$.
We usually choose α to be a regular grid of points; another possibility is that of choosing α to be a random sample of points within D. When A is a disc of radius t, this estimator corresponds to Ripley's estimate $\hat{p}(t)$ (Ripley, 1977). Thus the test set A is moved to a finite number of (generally overlapping) positions within the sampling window; the statistic $\hat{G}_x(A)$ is the proportion of times that the translated sets are empty of points of the pattern.

The second statistic used is analogous to $\hat{G}$, with an integral being used instead of a sum. We define

$$\tilde{G}_x(A) = \frac{1}{m(\beta)} \int_{\underset{\sim}{u} \in \beta} I(N_x(A_{\underset{\sim}{u}}) = 0) \, d\underset{\sim}{u} \qquad (1.7)$$

where $\beta = \{\underset{\sim}{u} \in \mathbb{R}^2 : A_{\underset{\sim}{u}} \subseteq D\}$. $\tilde{G}_x(A)$ is thus likely to be a more precise estimate of $G_x(A)$ than $\hat{G}_x(A)$; the test set is translated continuously rather than being moved to a discrete number of positions.

It is clear that $\hat{G}_x(A)$ and $\tilde{G}_x(A)$ are unbiased estimates of $G_x(A)$. Provided $P(I(N_x(A_u) + N_x(A_v) = 0) < (G_x(A))^2$ for $u \neq v$, $\tilde{G}_x(A)$ is likely to have smaller variance than $\hat{G}_x(A)$ and so should be preferred if it is possible to calculate it. Attention in this chapter is concentrated on the cases when A is a rectangle or an ellipse; $\hat{G}$ is used for rectangles, and $\tilde{G}$ for ellipses. The methods of computing these estimates are described in section 5.

Normally, in analysing a point pattern, we will want to consider test sets of several different shapes or sizes at the same time. Our statistic will thus be multivariate, of the form $(\hat{G}_x(A_1), \ldots, \hat{G}_x(A_n))$ or $(\tilde{G}_x(A_1), \ldots, \tilde{G}_x(A_n))$, where $A_1, \ldots, A_n$ are sets of different shapes or sizes. For example, we could have $A_i = U(t_i)$, $i = 1, \ldots, n$, where $t_1 < t_2 < \ldots < t_n$; this enables us to plot $\tilde{G}_x(U(t))$ against t and gives us an overall view of some of the properties of the pattern.

4. Asymptotic Results

Asymptotic distributional results can be proved for empty space statistics, under certain conditions, as the sampling window D increases to $\mathbb{R}^2$. These are based on the following result of Bulinskii and Zhurbenko (1976).

Lemma 1 (Bulinskii and Zhurbenko)

Let F(U) be a real-valued random function that is defined on the class $\mathcal{B}$ of all bounded Borel measurable subsets of $\mathbb{R}^n$ and is additive in the sense that $U' \cap U'' = \phi$ implies that

$$F(U' \cup U'') = F(U') + F(U'') .$$

Assume that $EF(U) = 0$, and write

$$\sigma^2(U) = \text{var } (F(U)).$$

For $\delta > 0$, define

$$C_{2+\delta}(U) = E|F(U)|^{2+\delta} .$$

Let $\alpha(U', U'') = \sup_{\substack{A \in \mathcal{U}(U') \\ B \in \mathcal{U}(U'')}} |P(A \cap B) - P(A)P(B)|$

where $\mathcal{U}(U)$ is the σ-algebra of events generated by the random variables F (C) for all measurable subsets C of the set U.

Define $\alpha^*(r) = \sup_{\rho(U', U'') \geq r} \alpha(U', U'')$,

where ρ is Euclidean distance in $\mathbb{R}^n$. Let $\{U_k\}$ be a sequence of subsets of $\mathbb{R}^n$.

For $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, write $\pi(a) = \{x \in \mathbb{R}^n : 0 < x_i \leq a_i, \; i = 1, \ldots, n\}$,

and for $m = (m_1, \ldots, m_n) \in \mathbb{Z}^n$, write $\pi_m(a) = \pi(a) + (m_1 a_1, \ldots, m_n a_n)$. For $\Lambda \subset \mathbb{R}^n$, we define $N_a^+(\Lambda)$ as the number of sets $\pi_m(a)$, $m \in \mathbb{Z}^n$, for which $\Lambda \cap \pi_m(a) \neq \phi$, and $N_a^-(\Lambda)$ as the number of $\pi_m(a)$, $m \in \mathbb{Z}^n$, such that $\pi_m(a) \subset \Lambda$. We say that the sets $\{U_k\}$ *converge to infinity in the sense of van Hove* if, for each $a \in \mathbb{R}^n$,

$$\lim_{k\to\infty} N_a^-(U_k) = \infty \quad \text{and} \quad \lim_{k\to\infty} N_a^-(U_k)/N_a^+(U_k) = 1 .$$

Suppose that the following conditions hold:

(A_2) For sets $U \in \mathcal{B}$,

$$\sigma^2(U) \asymp |U| \quad \text{as} \quad |U| \to \infty$$

(B_1) There exists $\delta > 0$ and constants C_o, ℓ_j^o, $j = 1, \ldots, n$ such that

$$C_{2+\delta}(\pi) \leq C_o$$

for all parallelepipeds π whose j^{th} edges have length ℓ_j^o, $j = 1, \ldots, n$.

$(D_2)'$ The sets $\{U_k\}$ converge to infinity in the sense of van Hove.

(C_4) $\alpha^*(r) = O(r^{-(n+\varepsilon)})$ for $\varepsilon > 2n/\delta$.

Then

$$\frac{F(U_k)}{\sigma(U_k)} \xrightarrow{\mathcal{D}} N(0,1) \quad \text{as} \quad k \to \infty .$$

We use the above lemma to prove the following asymptotic distributional result for empty space statistics.

Theorem A

Let A be a compact measurable subset of $\mathbb{R}^2$. Let X be a stationary point process in the plane. Define $U_k = [0,k] \times [0,k] \subseteq \mathbb{R}^2$, and

$$F_A(U) = \int_{\underset{\sim}{u} \in U} I(N_x(A_{\underset{\sim}{u}}) = 0)\, dm(\underset{\sim}{u}) - G_x(A)\, m(U) \qquad (1.8)$$

Let

$$\alpha_A(U', U'') = \sup_{\substack{A \in \mathscr{U}(U') \\ B \in \mathscr{U}(U'')}} |P(A \cap B) - P(A)P(B)|$$

where $\mathscr{U}(U)$ is the σ-algebra of events generated by the random variables $F_A(C)$ for all measurable subsets C of the set U.

Define $\alpha_A^*(r) = \sup_{\rho(U', U'') \geq r} \alpha_A(U', U'')$.

Let $G_k(A) = \dfrac{F_A(U_k)}{m(U_k)}$, $k = 1,2, \ldots$

Then, provided

$$\alpha_A^*(r) = O(r^{-(2+\eta)}) \text{ for some } \eta > 0 , \qquad (1.9)$$

we have

$$\{m(U_k)\}^{\frac{1}{2}} G_k(A) \xrightarrow{\mathscr{D}} N(0,\sigma^2) \quad \text{as} \quad k \to \infty$$

provided $\sigma^2 > 0$,

where

$$\sigma^2 = \int_{\underset{\sim}{u} \in \mathbb{R}^2} \{P(N_x(A_{\underset{\sim}{o}}) + N_x(A_{\underset{\sim}{u}}) = 0) - G_x(A)^2\}\, dm(\underset{\sim}{u}) \qquad (1.10)$$

Proof

The result follows from Lemma 1 above, with $n = 2$.

F_A is an additive real-valued random function. For $\delta > 0$, define $C_{2+\delta}(U)$ as in Lemma 1.

Now, $|F_A(U)| \leqslant m(U)$, so that $C_{2+\delta}(\pi) \leqslant 1$ for all parallelepipeds π all of whose edges are of length $\leqslant 1$. Thus condition (B_1) is satisfied. Define $f_x(y,x) = P(N_x(A_y) = N_x(A_z) = 0)$.

Then

$$\sigma^2(U_k) = \int_{y\in U_k} \int_{z\in U_k} (f_x(y,z) - G_x(A)^2)\, dydz,$$

and, by stationarity of X,

$$\frac{\sigma^2(U_k)}{m(U_k)} \to \sigma^2 \quad \text{as} \quad k \to \infty.$$

Thus, provided $\sigma^2 > 0$, condition (A_2) is satisfied.

Taking $\delta > 4/\eta$ ensures that condition (C_4) is satisfied. Condition $(D_2)'$ also holds, so that the result follows; the expression for σ^2 is easily verified. ∎

Note that here, $\{U_k\}$ can be replaced by any sequence of subsets converging to infinity in the sense of van Hove.

The above theorem can be applied in the following cases:

i) X is a Poisson process. In this case,

$$\alpha_A^*(r) = 0 \quad \text{for} \quad r > \text{diam}(A)$$

ii) X is a Poisson cluster process (see Chapter 0). In this case

$$\alpha_A^*(r) = 0 \quad \text{for} \quad r > r_o + \text{diam}(A) \ ,$$

where r_o is the diameter of the clusters.

iii) X is a Matérn hard core process of type 1 or 2, (described in Chapter 4, section 2). In this case,

$$\alpha_A^*(r) = 0 \quad \text{for } \tfrac{1}{2}r > R + \text{diam}(A) \ ,$$

where R is the diameter of the discs.

The above theorem gives asymptotic Normality of the empty space statistic $\tilde{G}$ under appropriate conditions. For the statistic $\hat{G}$ in the case where α is a square or rectangular grid (as is usually the case), asymptotic Normality is easily proved by using the corresponding result for the space $\mathbb{Z}^2$ (see the remark at the bottom of page 689 of Bulinskii and Zhurbenko (1976)).

The condition that σ^2 be positive is one that can be checked for any particular model for X; it certainly holds for all the models to which we will apply the theorem.

For the case of circular test set A, an explicit expression for the asymptotic variance σ^2 of $\sqrt{m(D)}\ (\tilde{G}_X(A) - G_X(A))$ can be obtained when X is a Poisson process. If A has radius t and X has intensity λ, then

$$\sigma^2_t = 4\pi t^2 e^{-2\lambda\pi t^2}\left[2\int_0^1 u \exp\{2\lambda t^2 (\cos^{-1}u - u\sqrt{1-u^2})\}\, du - 1\right] .$$

For given values of t and λ, σ^2_t can be calculated by numerical integration. A corresponding expression can be obtained for the asymptotic covariance σ_{st} of $\sqrt{m(D)}\ \tilde{G}_X(U(s))$ and $\sqrt{m(D)}\ \tilde{G}_X(U(t))$.

Analogous, though more complicated, expressions can be obtained for σ^2_t and σ_{st} when X is a Poisson cluster process.

As noted in section 3, we often use a multivariate empty space statistic; we can also prove the asymptotic normality of such statistics:

Theorem B

Let $A_1, \ldots, A_n$ be compact measurable subsets of $\mathbb{R}^2$. Let X be a stationary point process in the plane. Define U_k and $G_k(A_i)$, $1 \leqslant i \leqslant n$, as in Theorem A. Let $\alpha^*(r)$ be as in Theorem A. Then, if for $i = 1, \ldots, n$,

$$\alpha^*_{A_i}(r) = O(r^{-(2+\eta)}) \text{ for some } \eta > 0$$

we have

$$\{m(U_k)\}^{\frac{1}{2}} \begin{pmatrix} G_k(A_1) \\ \vdots \\ G_k(A_n) \end{pmatrix} \xrightarrow{\mathscr{D}} N_n(0, \Sigma) \quad \text{as } k \to \infty$$

provided $\sigma_{ii} > 0$, $i = 1, \ldots, n$,

where

$$\Sigma = (\sigma_{ij}) \ , \ 1 \leqslant i,j \leqslant n \ ,$$

and

$$\sigma_{ij} = \int_{\underset{\sim}{u} \in \mathbb{R}^2} \{P(N_x(A_{i\underset{\sim}{u}}) + N_x(A_{j\underset{\sim}{u}}) = 0) - G_x(A_i)G_x(A_j)\} \, dm(\underset{\sim}{u}) \qquad (1.11)$$

Proof

For $b_1, \ldots, b_n \in \mathbb{R}$, let $\underset{\sim}{b} = (b_1 \ \ldots \ b_n)$ and set

$$F_A(U) = \int_{\underset{\sim}{u} \in U} \sum_{i=1}^{n} b_i \, I(N_x(A_{i\underset{\sim}{u}}) = 0) \, dm(\underset{\sim}{u})$$

$$- \sum_{i=1}^{n} b_i \, G_x(A_i) \, m(U) \ .$$

Arguments directly analogous to those in the proof of Theorem A show that

$$m(U_k)^{\frac{1}{2}} (b_1 G_k(A_1) + \ldots + b_n G_k(A_n)) \xrightarrow{\mathcal{D}} N(0,\sigma_{\underset{\sim}{b}}^2) \text{ as } k \to \infty ,$$

where $\sigma_{\underset{\sim}{b}}^2 = b^T \Sigma b$. Since this is true for all $\underset{\sim}{b} \in \mathbb{R}^n$, $\underset{\sim}{b} \neq \underset{\sim}{0}$, this is enough to prove the required result. ∎

Theorem B can be compared with the result of Baddeley (1980), who proves a functional central limit theorem for $\hat{p}(t)$ - corresponding to the case where $A_i = U(t_i)$, $t < \ldots < t_n$. In fact, Baddeley's methods can easily be used for any shape of test set. Note, however, that his result is different from ours - it is proved by a different method, and our mixing condition (1.9), is much weaker than Baddeley's, which is that

$$\int_0^\infty \phi(t)\ t^5\ dt < \infty$$

where

$$\forall\ A \in \mathcal{U}(L_o),\ B \in \mathcal{U}(R_t),$$

$$|P(A \cap B) - P(A)P(B)| \leqslant \phi(t)P(A);$$

here, $L_o = \mathbb{R} \times (-\infty, 0]$ and $R_t = \mathbb{R} \times [t, \infty)$.

The asymptotic results stated in the above theorem can easily be adapted to multitype processes (see Chapter 3) and to 'sparse' rather than empty sets (see section 8). The methods of proof for sparse sets exactly follow those given above; the results for multitype processes are proved in Chapter 3.

5. Calculation of the estimates

In this section, we describe algorithms for calculating the empty space estimates $\hat{G}$ and $\tilde{G}$ in particular cases. Both these algorithms have been implemented in FORTRAN, and have been used in the examples that follow to calculate the estimates.

i) Calculation of $\hat{G}$ for a rectangular test set

Suppose that our sampling window D is rectangular, with sides of length s and t; we describe an algorithm for calculating $(\hat{G}_x(A_1), \ldots, \hat{G}_x(A_n))$, where A_i is a rectangle with sides of length $\frac{is}{p}$ and $\frac{it}{q}$, where $p,q \in \mathbb{N}$ and $i < \min(p,q)$.

The calculation is carried out by first placing a p×q rectangular grid over the sampling window D and noting which of the rectangles are empty of points of the pattern. All the subsequent calculations can be carried out using this information alone, by means of simple operations on the array recording whether or not each rectangle is empty.

For an $\frac{as}{p} \times \frac{bt}{q}$ rectangle, we use as our set α of possible positions the top left-hand (p-a+1) × (q-b+1) corner of the grid, and consider the test set positioned so that its top left-hand corner is at the relevant point of α (see Figure 1).

We use a 'logical' array E(i,j). Initially, E(i,j) is.TRUE.if and only if the grid cell with top left-hand corner at grid-point (i,j) is empty. There are two main operations involved in the calculation:

a) Operation H replaces E(i,j) by E(i,j).AND.E(i,j+1) for each permissible i and j. ('Horizontal addition')

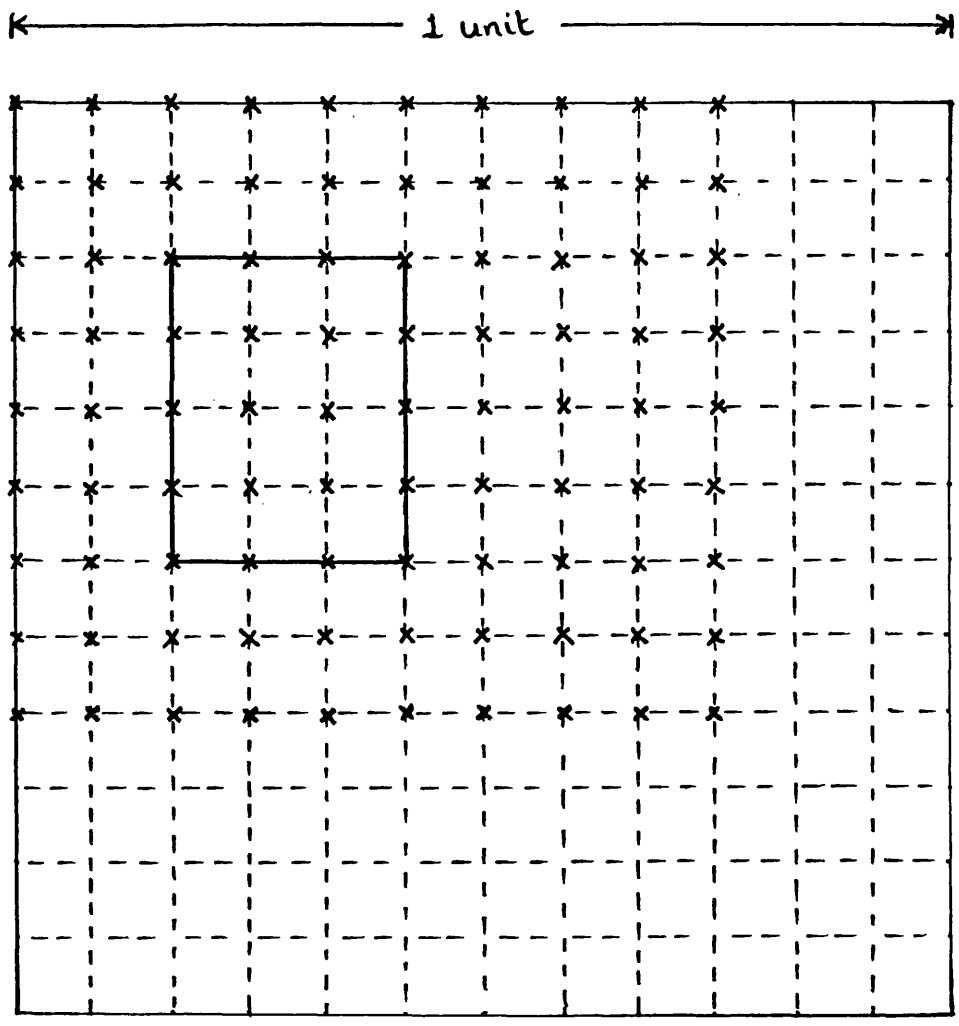


Figure 1 Calculation of the empty space statistic for a rectangular test set. The $\frac{3}{12} \times \frac{4}{12}$ rectangle shown is placed with its top left-hand corner at each of the positions marked by a cross.

b) Operation V replaces E(i,j) by E(i,j).AND.E(i+1,j)

for each permissible i and j. ('Vertical addition')

Thus, after a operations of type H and b operations of type V, the array variable E(i,j) ($i \leqslant p+1-a$, $j \leqslant q+1-b$) is .TRUE. if and only if the $\frac{as}{p} \times \frac{bt}{q}$ rectangle with top left-hand corner at the $(i,j)^{th}$ grid point is empty; thus a simple operation extracts $\hat{G}_x(A)$, where A is an $\frac{as}{p} \times \frac{bt}{q}$ rectangle. Performing the H and V operations in the correct sequence makes calculation of statistics of the form $(\hat{G}_x(A_1), \ldots, \hat{G}_x(A_n))$ straightforward. Similar ideas to these were used by Guild and Silverman (1978) to calculate their 'empty space' statistics.

ii) Calculation of $\tilde{G}$ for a circular or elliptical test set

We consider first the case of a circular test set of radius r. Suppose that our sampling window D is rectangular; we will need r to be less than half the length of the shorter side of D. Let C be the set of points u in D such that a closed disc of radius r centred at u contains no point of the sample. Then if A is a disc of radius r,

$$\tilde{G}_x(A) = \frac{m(C \cap J_r)}{(1-2r)^2}$$

where J_r is the rectangle whose sides lie distance r inside the sampling window D (see Figure 2).

We consider first the calculation of the area of the set C ; the algorithm for this calculation is due to Professor R. Sibson. We then discuss an edge-correction which enables us to calculate the area of $C \cap J_r$, and hence the unbiased estimate $\tilde{G}_x(A)$.

Note that C is the complement of the union of the open discs of radius r centred at the points of the pattern. We calculate the area

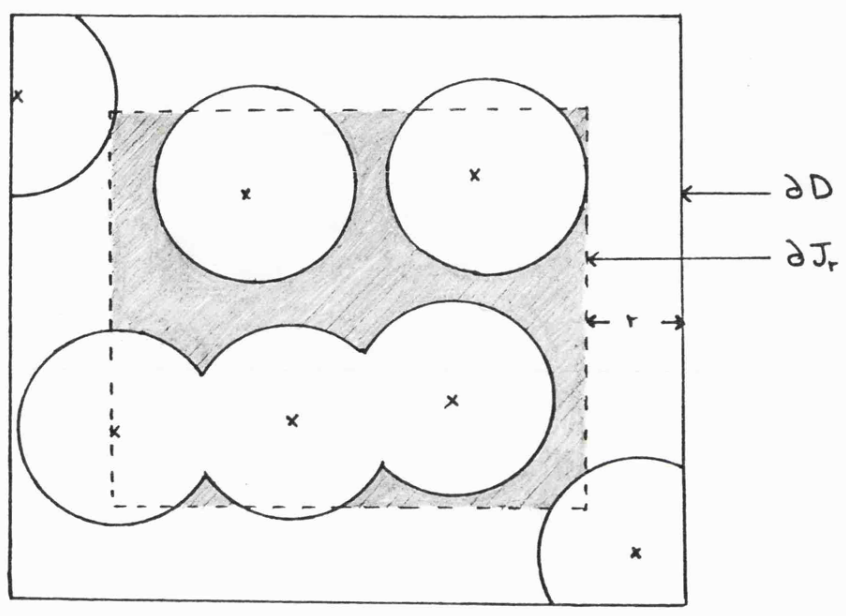


Figure 2 Calculation of $\tilde{G}(U(r))$. The shaded region is $C \cap J_r$.

of this union; this will provide us with the area of C. The obvious method of calculating the area of a union is by means of an inclusion/exclusion formula, but this quickly becomes impractical when intersections of many discs have to be handled. This occurs when the radius r approaches the spacing of the points of the pattern, and there is little empty space, which is the interesting case. Ripley (1977), in calculating his estimator $\hat{p}$, calculates the area by systematic or random test sampling within the window - a somewhat unsatisfactory procedure.

Our approach is a different one, and is based on the use of the Dirichlet tessellation. Green and Sibson (1978) have shown how to compute this construct efficiently for very large numbers of points. We first construct the tessellation defined by the pattern. Then, for a chosen value of r, we consider the intersection of the disc of radius r centred at a point of the pattern with the tile of that point. In general, the boundary of such a set will consist partly of circular arcs and partly of straight line segments (see Figure 3). The union of these sets is, to within a set of measure zero, the set in which we are interested, and they are disjoint. Calculating the area of such a set is tedious but straightforward; it is most easily accomplished by breaking the tile into triangles with a common vertex at the associated point of the pattern, and handling each such triangle separately. To calculate the total area, we sum first within each tile, and then over contributions from different tiles. Figure 3 illustrates the two stages of the decomposition.

The implementation of the above algorithm for calculating the area of the union of discs, written by Professor Sibson, also calculates the length of the boundary of the union, again by subdividing each tile into triangles. Knowledge of this boundary length makes it possible to

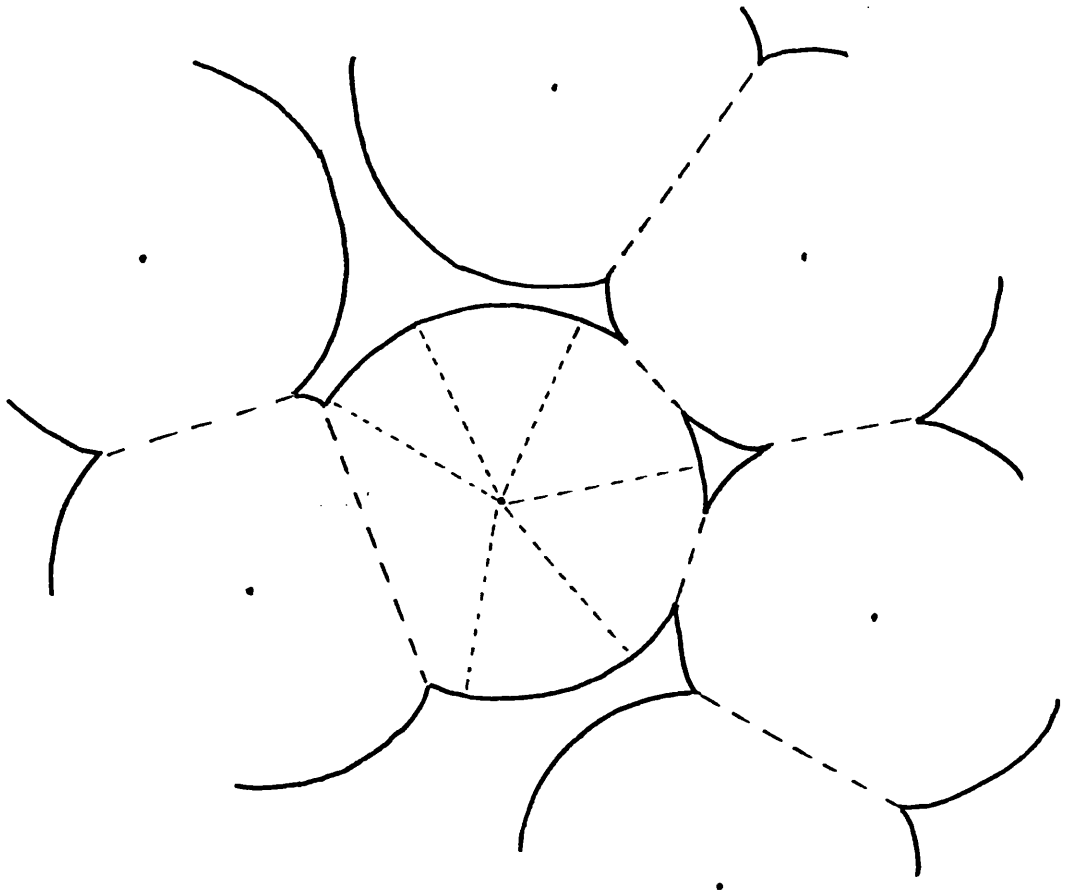


Figure 3 Decomposition of the union of the discs between and within tiles.

calculate the derivative of the area; this could be of value in interpolating smoothly between calculated values of the empty space statistic.

Calculating the edge-correction required to furnish us with the area of $C \cap J_r$ is, again, tedious but straightforward; we calculate the area of the intersection of the union of the discs with the smaller rectangle J_r (see Figure 4). For tiles lying wholly inside J_r, the calculations are as above; there is no contribution from tiles lying outside J_r, but tiles which intersect the border of J_r have to be considered separately. The contribution from such tiles is calculated by dividing parts of the tile into triangles with a common vertex at the corresponding point of the pattern; this time, some of the triangles will have a segment of the boundary of J_r as an edge. Some of the cases arising are shown in Figure 5. If the point of the pattern corresponding to the tile lies inside J_r, the contribution from that tile can be evaluated by direct summation of areas; otherwise, we have to difference the contributions from various triangles (see Figure 6). Different special cases have to be considered, but all the calculations are straightforward. The logic of the algorithm for computing the empty space statistic is illustrated in Figure 7.

Note that the tessellation needs to be calculated only once, and can then be interrogated using any number of different values of r.

We will occasionally wish to regard the rectangle D as a torus, so that points on opposite edges are considered to be near - this corresponds to imposing periodic boundary conditions. This situation arises particuarly when calculating empty space statistics for multitype patterns, as described in Chapter 3. To calculate the statistic in this case, we simply regard D as part of a grid of identical rectangles

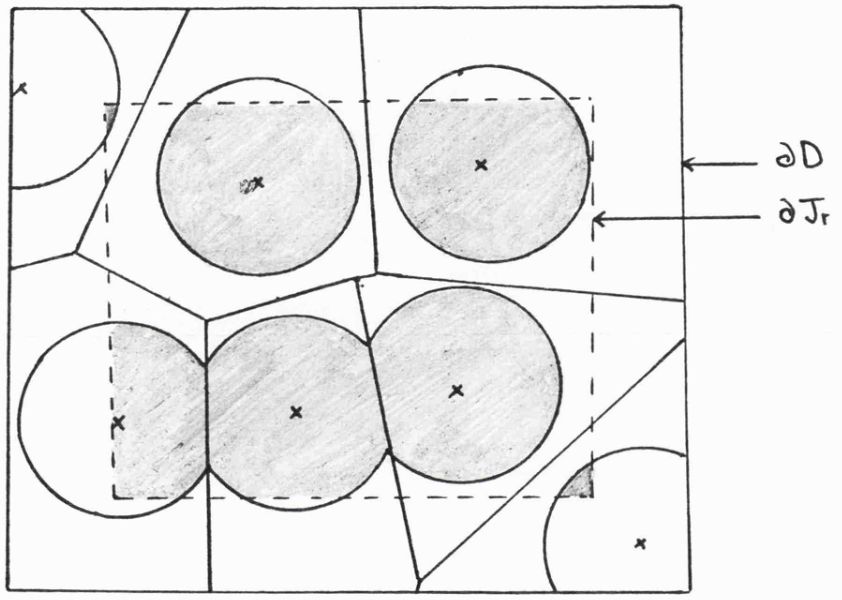


Figure 4 The shaded region is the intersection of the union of the discs with J_r. The tile boundaries are also shown.

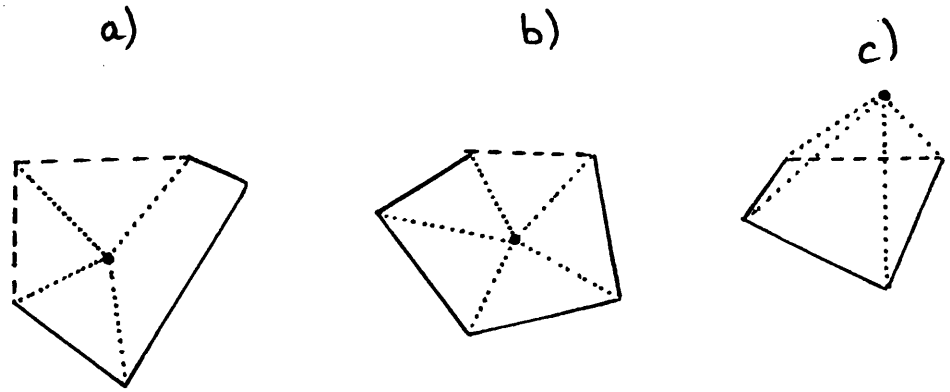


Figure 5 Some cases arising in the triangulation of J_r. The dashed lines are parts of ∂J_r, and the solid lines are tile boundaries. In case c), the generating point of the tile lies outside J_r.

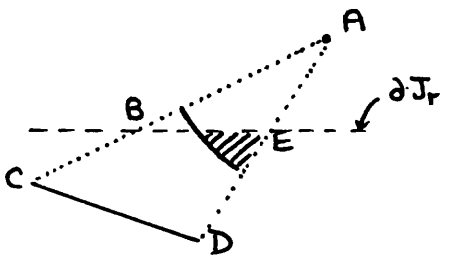


Figure 6 Contribution to the area from a triangle whose generating point lies outside J_r. The area is obtained by subtracting the contribution from triangle ABE from that from triangle ACD.

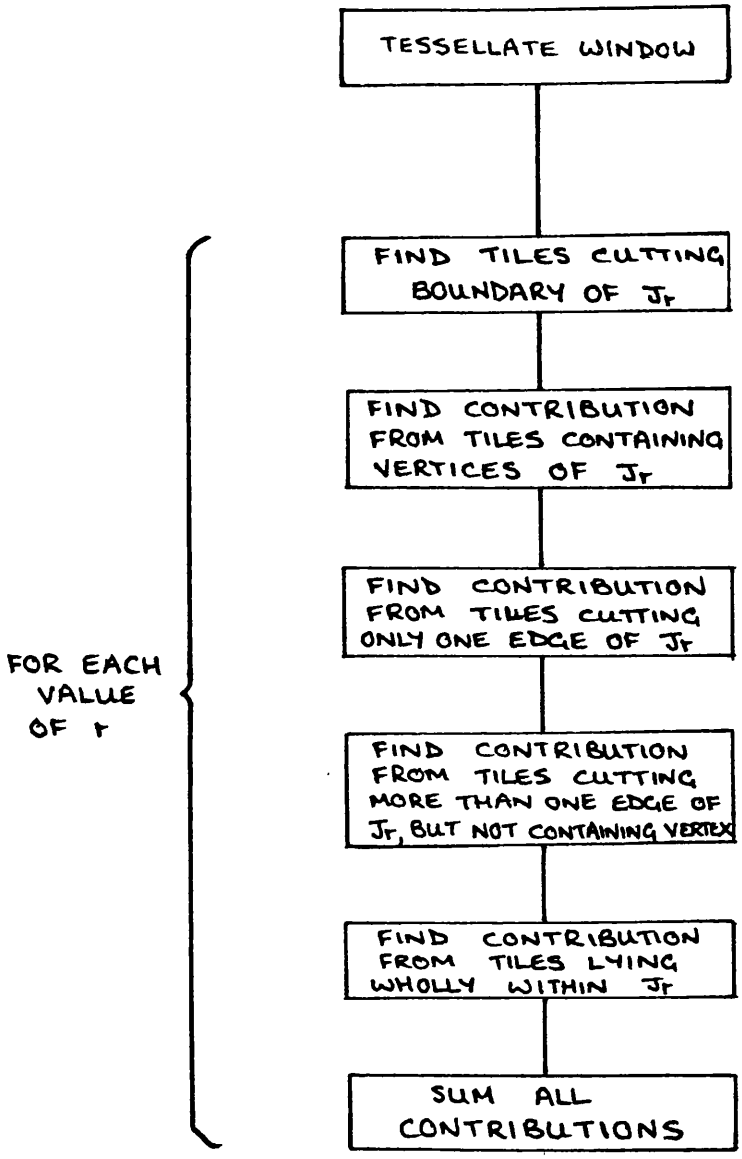


Figure 7 Logic of the algorithm for calculating $\tilde{G}(U(r))$.

(see Figure 8) and calculate the area of the intersection of the union with the inner rectangle; the method for doing this is the same as that for the edge-correction described above. In practice, we need only tessellate the rectangle D_r shown in Figure 8; this is a rectangle which gives a border of side r_{max} around D, where r_{max} is the largest radius of disc which we wish to consider.

To calculate the empty space statistics $\tilde{G}_x(A)$ for an elliptic test set A, we perform a suitable directionally non-uniform dilation of the original point pattern, and then treat the transformed pattern as a new data set to which a circular test set is to be applied. Division of the resulting area by the Jacobian of the dilation recovers the area associated with the original pattern. Using an elliptic, rather than circular, test set could be of value in investigating anisotropy in the data. (Compare Guild and Silverman (1978), who used rectangular test sets to investigate anisotropy).

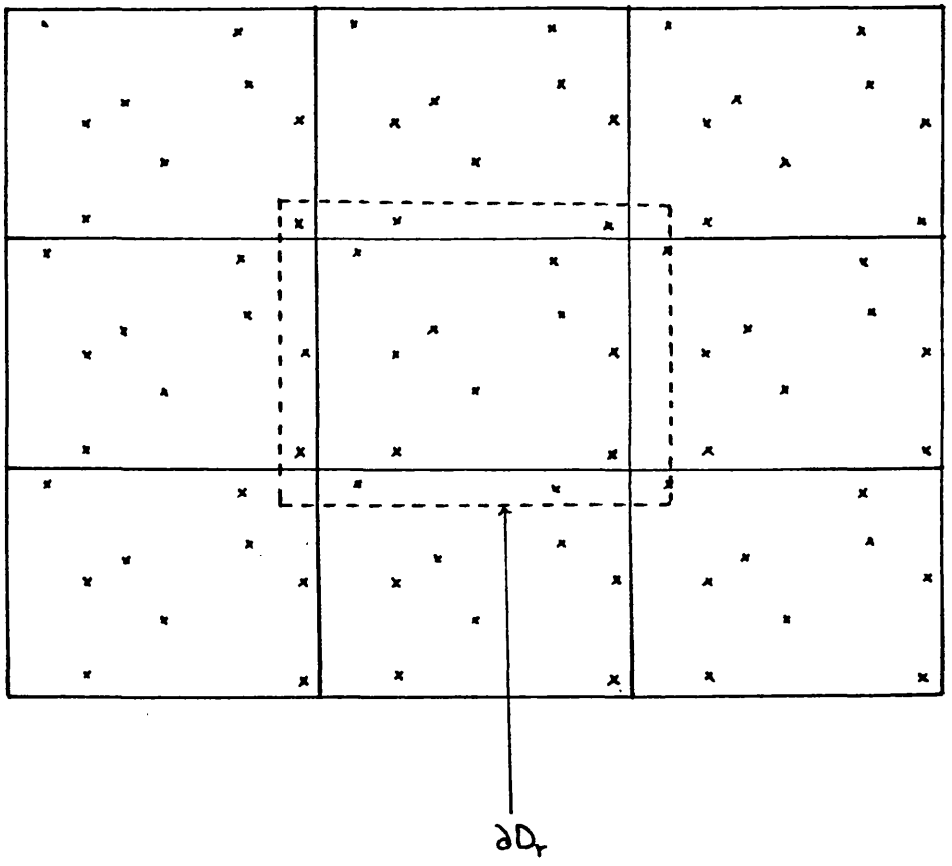


Figure 8 Grid of rectangles used for calculating the empty space statistic on a torus. The central rectangle is the sampling window D.

6. Using the Empty Space Statistic

In this section, the use in practice of empty space statistics is described, the methods being illustrated by application to two point patterns, taken from the Lansing Woods data of Prof. D.J. Gerrard. The first tree pattern consists of the locations of 135 black oaks in a 924 ft. by 924 ft. region of central Michigan;the second is a pattern of 514 maples in the same region. The data are described more fully in Gerrard (1969); for other analyses of some of the Lansing Woods data, see Besag (1977), Diggle (1977 and 1980), and Cox and Lewis (1976).

i) Black Oaks

The data are shown in Figure 9; inspection of the pattern suggests that it is clustered. The pattern was first analysed using square test sets, the estimates being calcuated from a 100 × 100 grid placed over the sampling window; the length of the side of the window is taken as the unit of measurement in this and subsequent analyses of Lansing Woods data. The results of this analysis are shown in Figure 10; as expected, the pattern is shown to be highly clustered. Use of a circular test set yielded similar results, shown in Figure 11. Besag (1977) and Cox and Lewis (1976), using other methods, also found evidence of clustering in this pattern.

The nature of the empty space curves suggests the possibility of attempting to fit some kind of a Poisson cluster process to the pattern; this is done, using other techniques, in section 5 of Chapter 2.

ii) Maples

The pattern is shown in Figure 12. Here, we have many more points, so would expect stronger results. The pattern again appears clustered. The data were analysed using square and circular test sets, and the results are shown in Figures 13 and 14. The clustering of the pattern

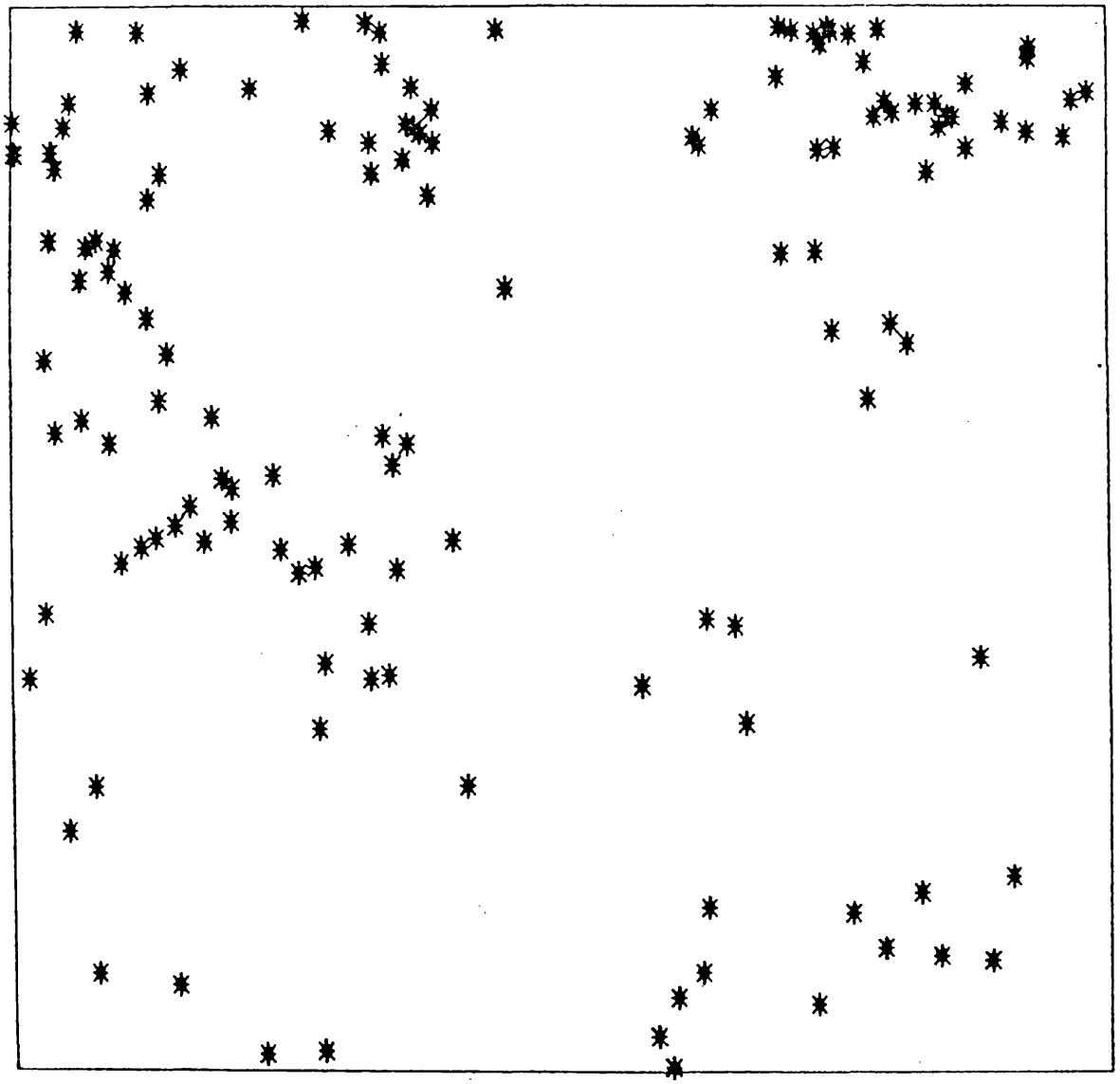


Figure 9 The positions of 135 black oaks in a 924 ft by 924 ft region of Lansing Woods.

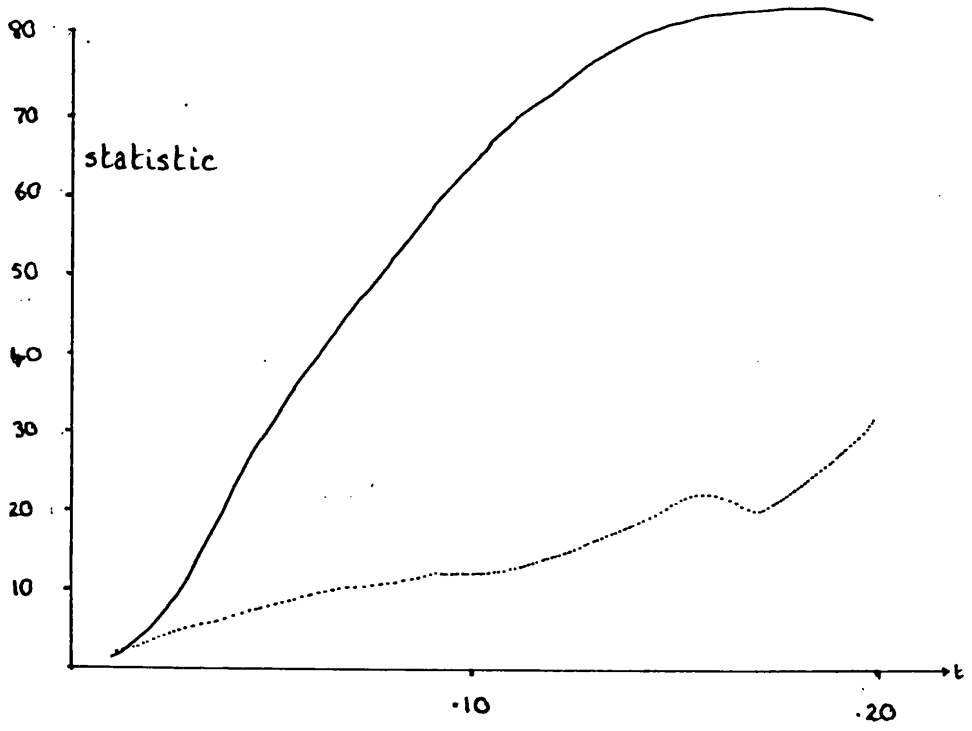


Figure 10 Empty space for the black oaks, using a square test set. The statistic plotted is $\{\log (\hat{G}(S(t))) + 135t^2\}/t^2$. The dotted line is a simulated upper 95% confidence band for a Poisson process conditioned to have 135 points.

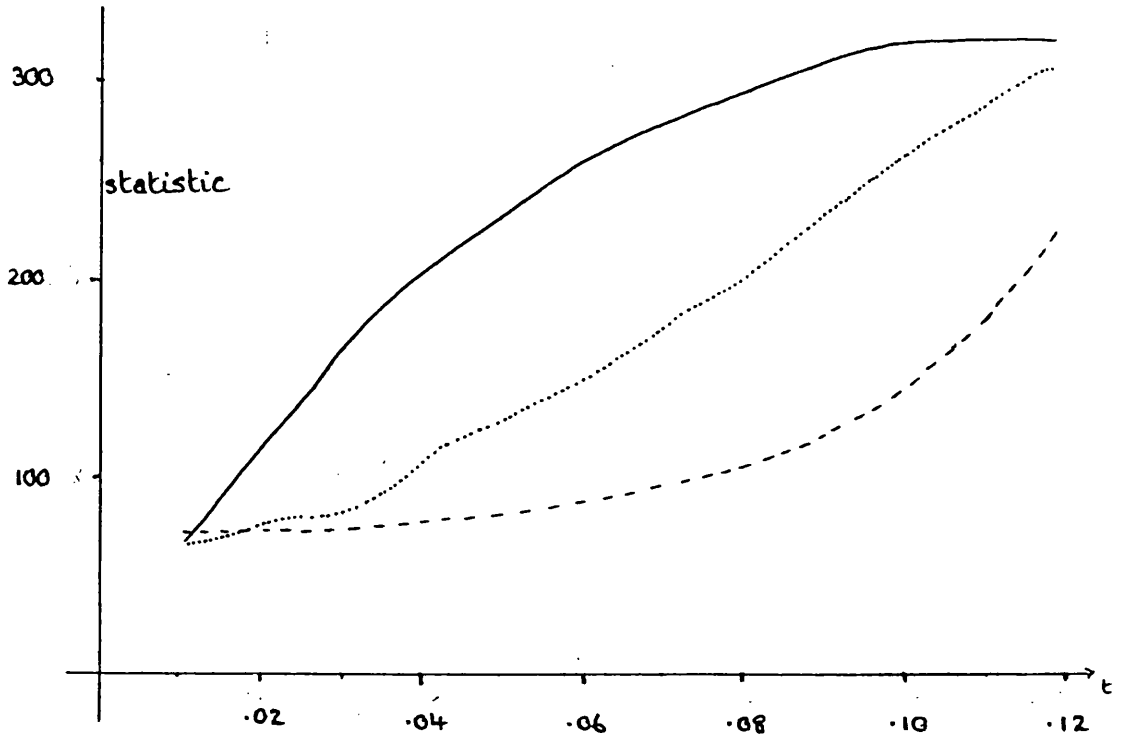


Figure 11 Empty space for the black oaks, using a circular test set. The statistic plotted is $\{\log (\tilde{G}(U(t))) + 135\pi t^2\}/t^2$. The dashed line is the upper asymptotic 95% confidence band for a Poisson process of corresponding intensity; the dotted line is a simulated 95% confidence band for a Poisson process, obtained from 99 simulations.

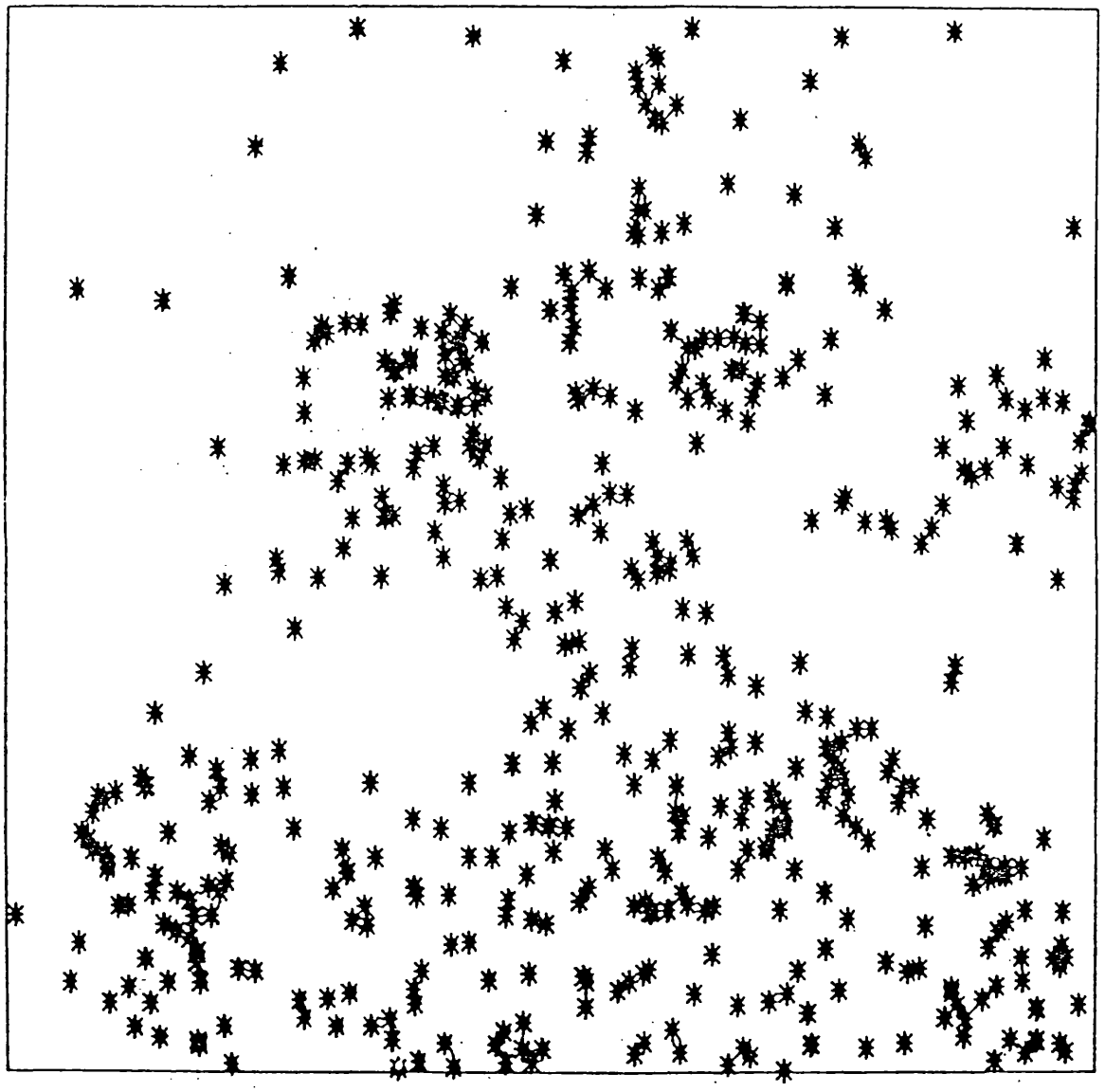


Figure 12 The positions of 514 maples in a 924 ft by 924 ft region of Lansing Woods.

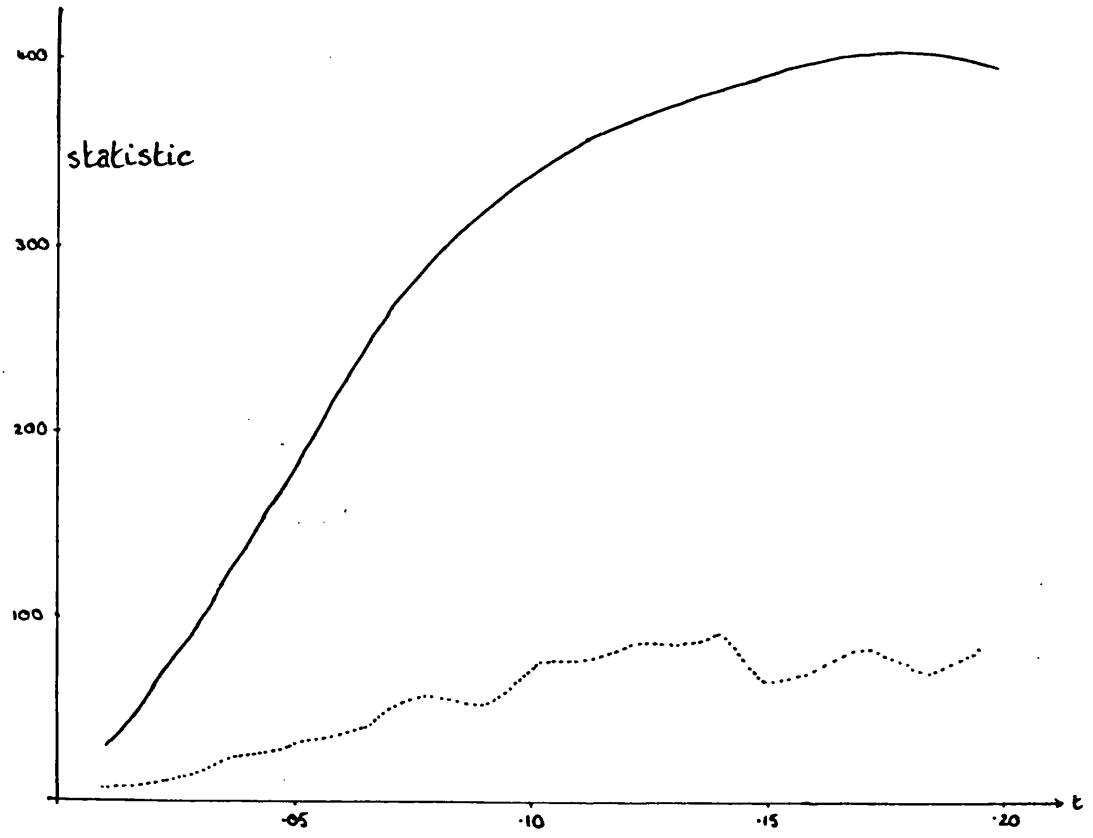


Figure 13 Empty space for the maples using a square test set. The statistic plotted is $\{\log(\hat{G}(S(t))) + 514\ t^2\}/t^2$. The dotted line is a simulated upper 95% confidence band for a Poisson process conditioned to have 514 points.

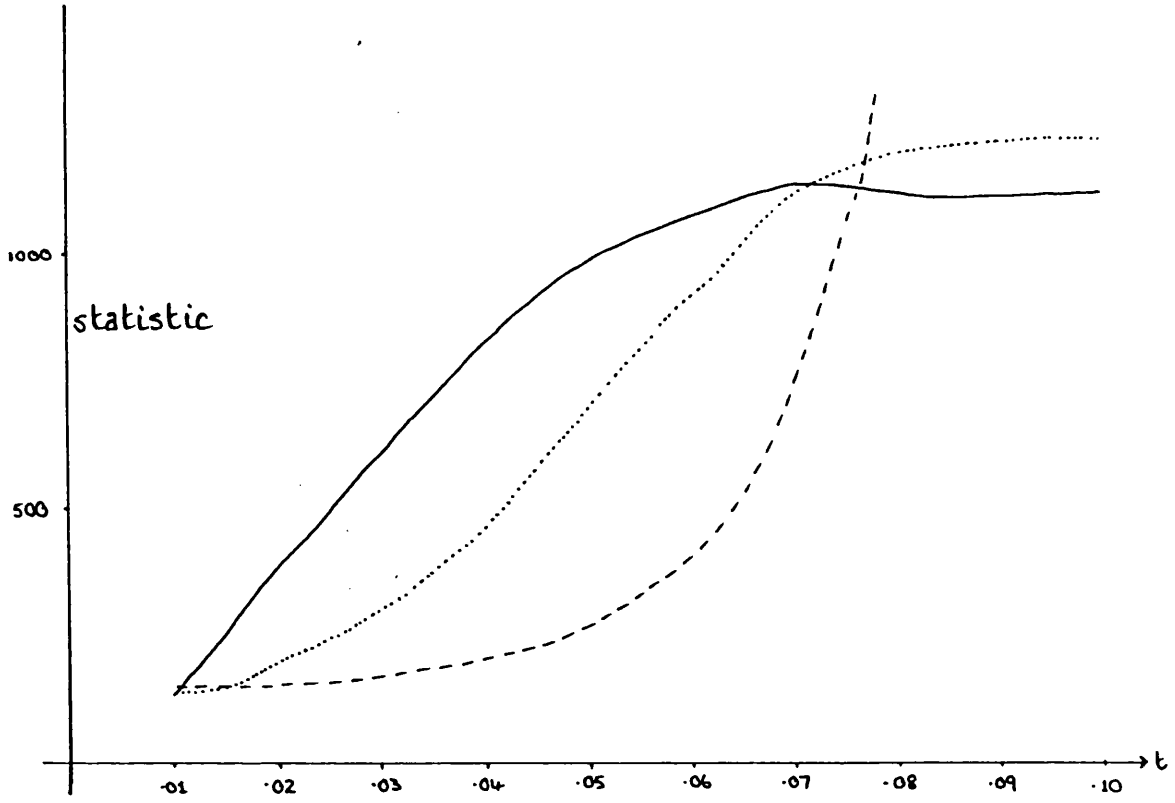


Figure 14 Empty space for the maples, using a circular test set. The statistic plotted is $\{\log(\tilde{G}(U(t))) + 514\pi t^2\}/t^2$. The dashed line is the upper asymptotic 95% confidence band for a Poisson process of corresponding intensity; the dotted line is a simulated 95% confidence band for a Poisson process, obtained from 99 simulations.

is again apparent. This agrees with the results of Besag (1977), Diggle (1980) and Cox and Lewis (1976). A Poisson cluster process is fitted to this pattern in Chapter 2, section 5.

7. Power of tests of randomness based on empty space

i) Theoretical results

For models where it is known that the empty space statistic is asymptotically normally distributed, some theoretical progress is possible when considering the power of tests of randomness based on empty space. Suppose we have a realisation x of a point process X, and that we wish to test the null hypothesis.

H_o: X is a Poisson process of parameter λ against the alternative hypothesis.

H_1: X is a process of type T, with parameter $\underset{\sim}{\theta}$. (Here, 'type T' could mean, for example, a Poisson-Poisson cluster process).

Let $A \subseteq \mathbb{R}^2$, and suppose that, for sampling window D, $\dfrac{\tilde{G}_x(A) - \mu_o}{m(D)^{\frac{1}{2}}}$ is asymptotically distributed as $N(0,\sigma_o^2)$ under H_o and that $\dfrac{\tilde{G}_x(A) - \mu_1}{m(D)^{\frac{1}{2}}}$ is asymptotically distributed as $N(0,\sigma_1^2)$ under H_1.

Suppose, for example, that $\mu_1 > \mu_o$ and that we use the asymptotic normality of $\tilde{G}_x(A)$ to provide an approximate one-sided size α test of H_o. Then the power of such a test will be approximately

$$1 - \Phi\left(\frac{\sigma_o}{\sigma_1}\ \Phi^{-1}\ (1 - \alpha) + \frac{\mu_o - \mu_1}{\sigma_1\, m(D)^{\frac{1}{2}}}\right) .$$

For Poisson-Poisson cluster processes (described in Chapter O), we can calculate μ_1 and σ_1 by numerical integration for a particular A and for given values of the clustering parameters. This gives us an approximate expression for the power of tests of randomness based on $\tilde{G}_x(A)$. We are more likely to want to use a multivariate statistic of the form $(\tilde{G}_x(A_1), \ldots, \tilde{G}_x(A_n))$ to test for randomness of the point pattern x. While this statistic will be asymptotically distributed

as a multivariate normal distribution for Poisson processes or Poisson cluster processes, and it is possible to calculate the variance/covariance matrix, calculating the probability of rejecting the null hypothesis becomes complicated, and it is more realistic to investigate the power of such tests by simulation methods; this is done in the remainder of this section.

ii) Simulation results

Ripley (1979b) carried out a simulation study of the power of tests based on distance methods. In this section, the results of a similar study for the empty space statistic are described. Ripley used Strauss processes and Matérn cluster processes (described in Chapter 0) as his alternatives to Poisson processes. As was noted in section 1, we would expect empty space methods to be able to detect clustering. 'Regular' alternatives were not considered; since the likelihood function for Strauss processes depends only on the inter-point distances, second order methods will be best for detecting such departures from Poisson, as the inter-point distances are a sufficient statistic.

The statistic used in the tests was $(\tilde{G}(t_i), t_i = 0.01 \times i, i = 1, \ldots, 20)$. The 5 per cent significance levels were obtained by Monte Carlo methods using 100 simulations of a binomial process; following Kelly (1977)'s suggestion, the number of times $\tilde{G}$ touches the envelope obtained from the simulations was used as a test statistic to obtain one-sided tests of size 5 per cent. The results of the simulations are shown in Table 1.

$\tilde{G}$ gives similar power to $\hat{K}$ for very clustered 100-point patterns, and is comparable with $\hat{K}$ for some other patterns; otherwise,

$\hat{K}$ is considerably better than the empty space statistic. The results of this section can also be compared with those of Diggle (1979), who used a different type of cluster process when comparing the performance of various statistics, including the empty space-like statistic d_x.

TABLE 1

Estimated power in per cent against a Matérn cluster process with mean number of clusters N_c and radius R. The corresponding figures for Ripley's L_m (taken from Ripley, 1979b) are included for comparison.

			POWER	
N	N_c	R	$\tilde{G}$	L_m
25	12	0.1	54	70
25	12	0.2	26	36
25	8	0.2	36	63
25	5	0.25	60	65
100	50	0.1	82	78
100	25	0.15	91	89
100	25	0.25	39	47
100	10	0.5	15	23

8. Generalisation : sparse sets

Thus far, we have discussed the use of empty sets in the analysis of spatial patterns. An obvious generalisation of the methods described above is to relax the condition that the test set be completely empty, and look instead for sparsely-populated sets. Formally, we can define a set A to be sparse if $N_X(A) \leqslant S$ for some given $S \in \mathbb{N}$. The choice of S here may depend on the size of A and on the intensity of the process X. We can then define the sparse set probability $G_X^S(A)$ by

$$G_X^S(A) = P(N_X(A) \leqslant S),$$

and, analogously to (1.6) and (1.7), define unbiased estimators $\hat{G}_X^S(A)$ and $\tilde{G}_X^S(A)$ by

$$\hat{G}_X^S(A) = \frac{1}{|\alpha|} \sum_{\underset{\sim}{u} \in \alpha} I(N_X(A_{\underset{\sim}{u}}) \leqslant S) \qquad (1.12)$$

and

$$G_X^S(A) = \frac{1}{m(\beta)} \int_{\underset{\sim}{u} \in \beta} I(N_X(A_{\underset{\sim}{u}}) \leqslant S)\, d\underset{\sim}{u} \; ; \qquad (1.13)$$

here, the notation is as in section 3. Note that empty space is simply the case S = 0. Analogous asymptotic results to those proved in section 4 hold; the proofs will go through in exactly the same way. Sparse sets can be used in ways similar to those described in section 6 for model-fitting or hypothesis testing. The usefulness of sparse sets remains an area for further study; such methods could be of particular use when some of the point positions may have been inaccurately recorded, or when a few spurious points may be present

(see, for example, Guild and Silverman (1978)). From the computational point of view, the use of rectangular, rather than circular test sets would appear to be best suited to 'sparse set' methods, so that estimates of the form $\hat{G}_x^S(A)$ would be used. The Dirichlet tessellation can be used to calculate $\tilde{G}_x^S(A)$ for a circular test set A, but, even for S = 1, such a calculation would be time-consuming.

9. Discussion : the usefulness of empty space

In the preceding sections, the use of empty space methods for analysing spatial point patterns has been described; such methods provide a useful alternative to distance-based tecnhiques. Empty space statistics furnish a different kind of information from that provided by inter-point distances, and plotting $\hat{G}(U(t))$ against t gives an informative picture of some of the properties of a point pattern.

While it is possible to use empty space on its own for model-fitting or hypothesis testing, the results of section 7 and of Diggle (1979) suggest that distance-based methods are better in this respect. This is possibly because of the type of models considered as alternatives to randomness in these studies; most of the commonly-used models are, at least in part, distance-based, and it might be possible to construct realistic models based on the idea of empty space. In any event, empty space methods for single-type patterns would seem to be most useful as a source of complementary information to that provided by distance methods. $\tilde{G}$ and $\hat{K}$, for example, could be used together as exploratory tools in the analysis of a point pattern, or for the purposes of model fitting. In the latter case, $\hat{K}$ could be used to fit a parameter for a certain model, and $\tilde{G}$ could then be used to assess the goodness of the fit provided. This kind of approach is used in section 5 of Chapter 2 to fit Poisson cluster processes; the advantages of such a technique are discussed in that section.

Attention has been confined in this chapter to the use of empty space methods for analysing single-type point patterns. Empty space statistics can be very effective in the analysis of multitype patterns;

here again they are used alongside statistics based on inter-point distances. The use of empty space for multitype patterns is the subject of Chapter 3.

CHAPTER TWO

ESTIMATION OF THE SECOND MOMENT DENSITY USING HISTOGRAMS

1. Introduction

Ripley's second moment measure K(t) (Ripley, 1977), being a cumulative function, is difficult to interpret, especially visually, and the second moment density, g(t) (Ripley, 1977), is a more natural, if more difficult, function to estimate. Cox (1965) used a histogram estimate of the intensity function of a stationary point process on the line, and Peebles (1974b; see also 1974a) also used histograms to estimate the two-point correlation function w (interpreted by Ripley (1977) as $g/\lambda^2 - 1$) for three-dimensional astronomical data. In this chapter, we discuss the use of histograms in estimating the second moment density function g(t) for a stationary point process in the plane. Some theoretical properties of such estimates are given in section 3 and the use of histograms is illustrated by means of some examples in section 4. Besag (in the discussion of Ripley (1977)) has noted that in the modelling phase, a smoothed estimate of Peebles' w-function may be more informative than $\hat{K}$ or $\sqrt{\hat{K}}$; in section 5 of this chapter, we discuss and illustrate an ad hoc method of fitting a certain type of Poisson cluster process, using histogram estimates.

There are two different problems to which the use of histogram estimates can be applied:

i) Estimating g(t) for a given model. In this case, the model can be simulated a number of times, and the histograms obtained from each realisation can be averaged to produce the final estimate.

ii) Estimating g(t) for a particular dataset. Here, the histogram may oscillate considerably, so that some smoothing of the curve may be necessary in order to obtain a reasonable estimate of g(t). Such smoothing is described in section 2.

2. The Method

Under suitable conditions, given in the Appendix of Ripley (1977), the second moment density g(t) is related to the second moment measure K(t) by

$$g(t) = \frac{\lambda^2}{2\pi t} \frac{dK}{dt} \qquad (2.1)$$

where λ the intensity of the process.

Glass and Tobler (1971), in analysing a particular point pattern, plotted a histogram estimate of $g(t)/2\pi t$, using an edge correction similar to Ripley's (1977). We estimate $g(t)/\lambda^2$; our estimator is based on the approximately unbiased estimator $\hat{K}(t)$ of K(t) developed by Ripley (1977), and is obtained by differencing $\hat{K}$.

Suppose that we have a pattern x of N points in a finite region of area A, and that we wish to estimate the function g(t) for $c < t \leqslant d$. We form a histogram estimate of g(t) as follows. Form a partition $c = a_1 < a_2 < \ldots < a_{m+1} = d$ of the interval [c,d]. For $t \in (a_i, a_{i+1}]$, we estimate g(t) by

$$\hat{g}(t) = \frac{N(N-1)}{A^2} \times \frac{1}{2\pi(a_i + a_{i+1})/2} \times \frac{\hat{K}(a_{i+1}) - \hat{K}(a_i)}{a_{i+1} - a_i}$$

$$= \frac{N(N-1)}{A^2} \times \frac{\hat{K}(a_{i+1}) - \hat{K}(a_i)}{\pi(a_{i+1}^2 - a_i^2)} \qquad (2.2)$$

It is more convenient to estimate the normalised second moment density $g(t)/\lambda^2$; for $t \in (a_i, a_{i+1}]$, we estimate this function by

$$G_i = \frac{\hat{K}(a_{i+1}) - \hat{K}(a_i)}{\pi(a_{i+1}^2 - a_i^2)} \qquad (2.3)$$

Our histogram estimate can thus be summarised as $\{G_1, \ldots, G_m\}$. We normally choose the $\{a_i\}$ so that $a_i = \left(\frac{m - i + 1}{m}\right) c + \left(\frac{i - 1}{m}\right) d$, i.e. so that the bars of the histogram have constant width $b = \frac{d - c}{m}$; alternatively, we may, for reasons explained in section 3, wish to choose the $\{a_i\}$ so that $a_{i+1}^2 - a_i^2$ is constant. We will assume in the remainder of the chapter that the $\{a_i\}$ have been chosen in one of these two ways. The choice of the number, m, of bars in the histogram will depend on the effect desired in displaying the histogram, and on the precision required.

It was noted in section 1 that, when estimating $g(t)/\lambda^2$ from a single point pattern, smoothing of the histogram may be necessary in order to obtain a satisfactory curve. Figure 1 shows a histogram estimate of $g(t)/\lambda^2$ for a pattern of 200 points in the unit square using 75 bars between t = 0 and t = $3\sqrt{2}/8$; the histogram is plotted for $0 \leqslant t \leqslant 0.45$. The graph oscillates considerably, and will need to be smoothed before a pleasing pictorial representation is obtained. We smooth the histogram using simple weights based on the symmetric Binomial distribution: for S even, define the smoothed value $G_k^{(S)}$ at the k^{th} bar, for $\frac{S}{2} + 1 \leqslant k \leqslant m - \frac{S}{2}$, by

$$G_k^{(S)} = \frac{1}{2^S} \sum_{j=k-S/2}^{k+S/2} \binom{S}{\frac{S}{2} + k - j} G_j \qquad (2.4)$$

For $1 \leqslant k \leqslant S/2$ or $m - \frac{S}{2} - 1 \leqslant k \leqslant m$, we define $G_k^{(S)}$ in such a way that all the weight of the histogram will be redistributed in the smoothing:

$$G_k^{(S)} = \frac{1}{2^S} \sum_{j=j_o}^{j_1} \binom{S}{\frac{S}{2} + k - j} G_j + \left\{1 - \sum_{j=j_o}^{j_1} \frac{1}{2^S} \binom{S}{\frac{S}{2} + k - j}\right\} G_k$$

(2.5)

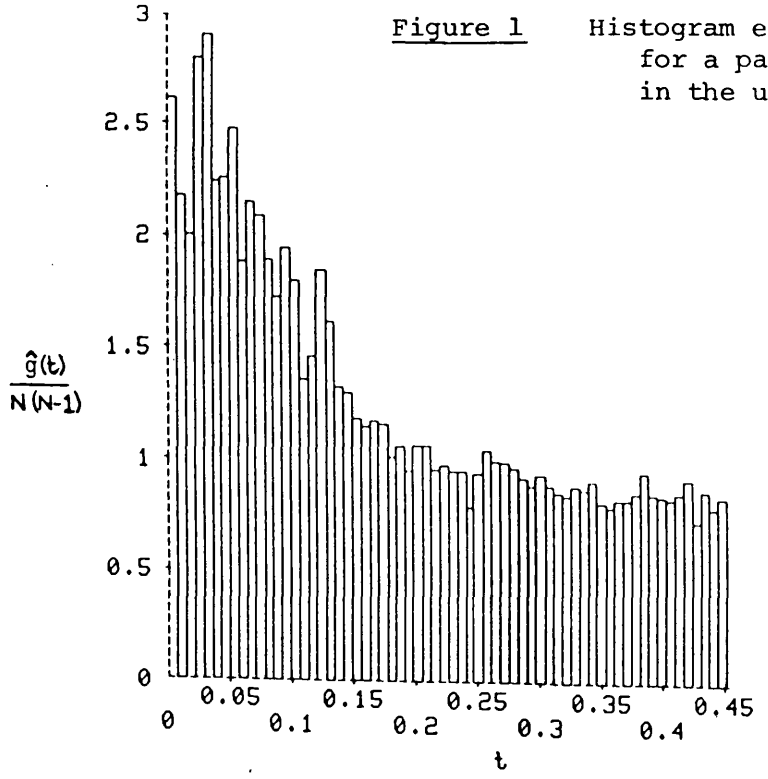


Figure 1 Histogram estimate of $g(t)/\lambda^2$ for a pattern of 200 points in the unit square.

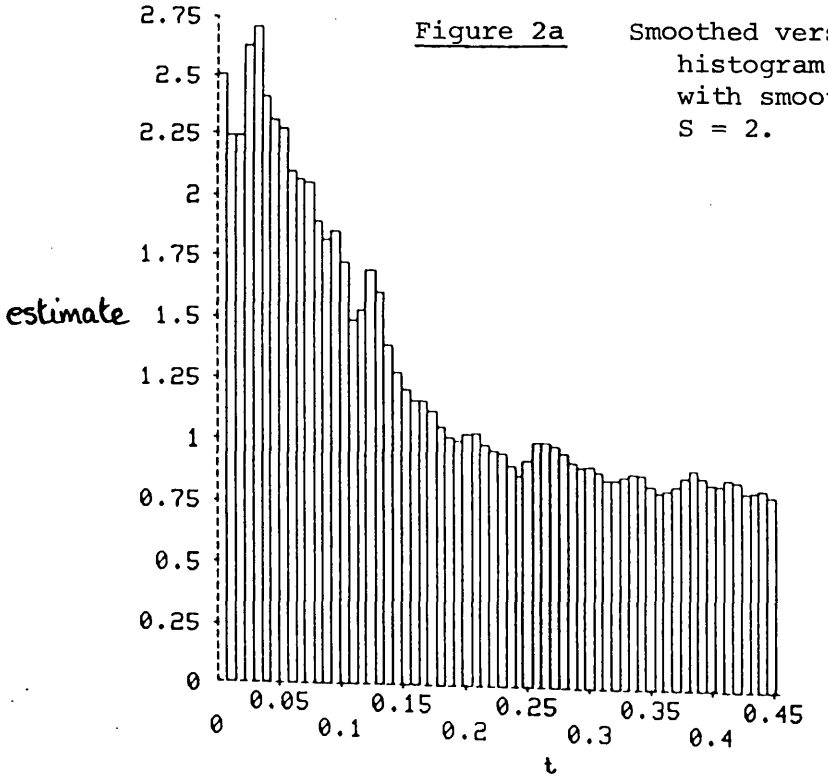


Figure 2a Smoothed version of the histogram of Figure 1, with smoothing parameter S = 2.

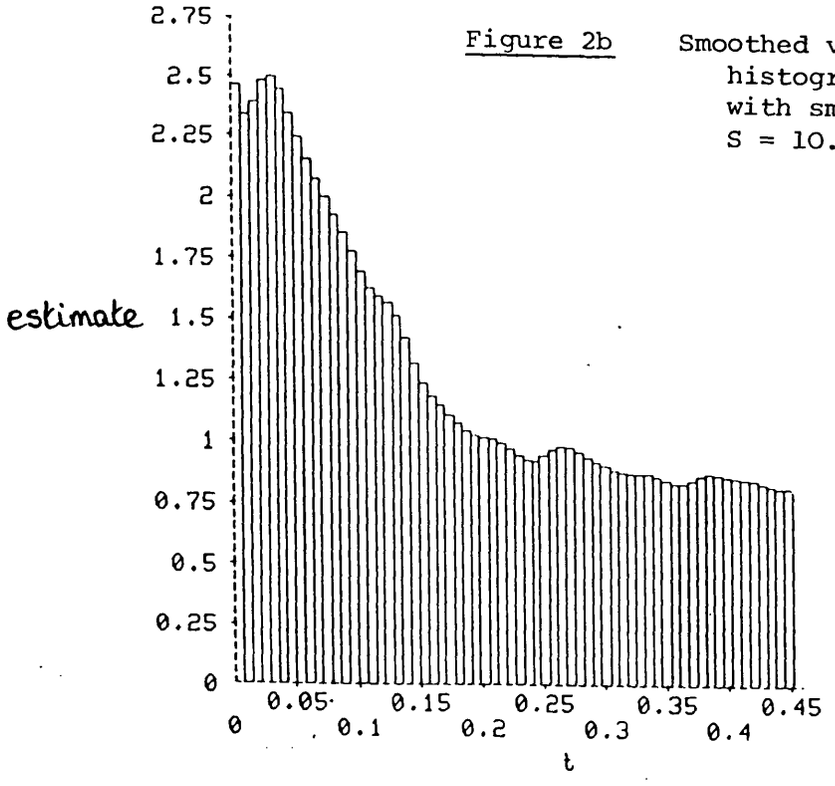


Figure 2b Smoothed version of the histogram of Figure 1, with smoothing parameter S = 10.

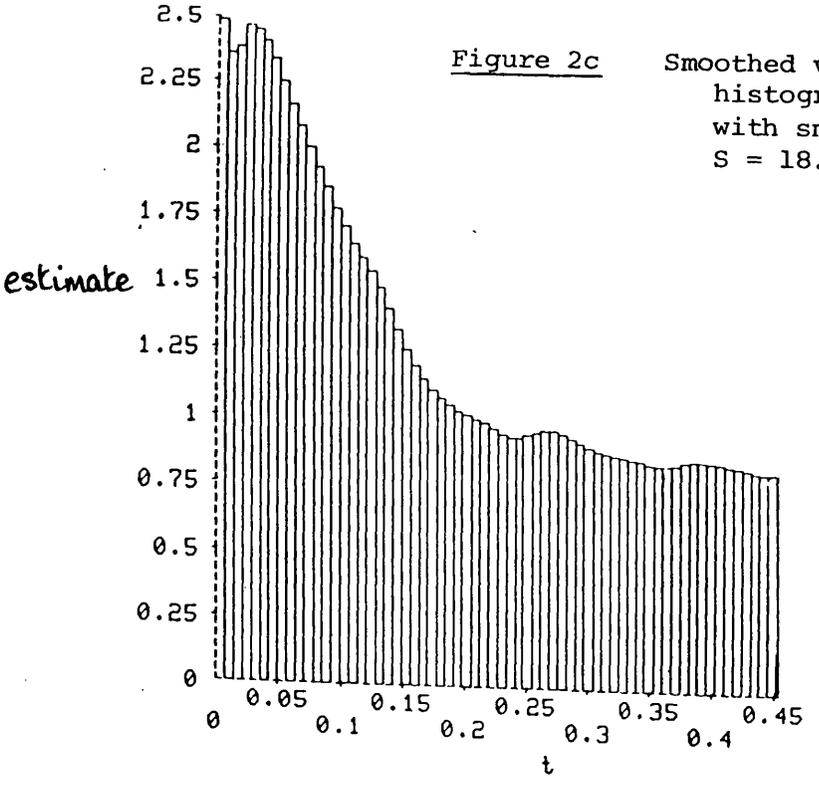


Figure 2c Smoothed version of the histogram of Figure 1, with smoothing parameter S = 18.

where $j_o = \max(1, k - S/2)$ and $j_1 = \min(m, k + S/2)$. Figure 2 shows the smoothed versions of the histogram of Figure 1, for various values of the smoothing parameter S. A satisfactorily smooth curve is obtained for S = 10. The choice of the smoothing parameter is discussed in the next section.

3. Some Theoretical Properties of the Histogram Estimates

i) Moments

Ripley (1977) notes that for a pattern of N points independently and uniformly distributed in the sampling region D, and

$$t \leqslant t_o = \inf\{s: \exists x \epsilon D,\ \partial Z_s(x) \cap D = \emptyset\},$$

$E(\hat{K}(t)) = \pi t^2$. Thus, for such a process, $EG_i = 1$ for each i.

If we assume that the points are independently and uniformly distributed on a rectangular torus, rather than in a rectangle, so that we use the unweighted empirical distribution function of the inter-point distances, rather than $\hat{K}$, it is easily shown that

$$\text{var}(G_i) = \frac{2}{N(N-1)} \left\{ \frac{1}{\pi(a_{i+1}^2 - a_i^2)} - 1 \right\} \qquad (2.6)$$

and

$$\text{cov}\ (G_i, G_j) = \frac{-2}{N(N-1)} \quad , \quad i \neq j. \qquad (2.7)$$

Making this assumption corresponds to imposing periodic boundary conditions on the pattern (see Ripley (1979b)). This simplifying assumption enables us to gain insight into the behaviour of the histogram for independently and uniformly distributed points (or, equivalently, for a Poisson process, conditional on the number N of points lying in the sampling window). For values of t that are small compared with the sides of the rectangle, there should be little difference between $\hat{K}$ and the empirical distribution function of the inter-point distances. Expression (2.6) shows that, if we have bars of constant width, then the variance of G_i will be greater for small values of i then for larger ones. Choosing the partition so that

$a_{i+1}^2 - a_i^2$ is constant stabilises the variance under the Poisson assumption. Using such bars of "constant t^2-width" has, however, the disadvantage of compressing that part of the histogram corresponding to short distances into a small interval - it is often the short distances that are of interest.

Obtaining expressions for the moments of histogram estimates for models other than the Poisson process is difficult. In general, the bias of G_i as an estimator of $g(t)/\lambda^2$, $t \in (a_i, a_{i+1}]$, will consist of three components:

a) a contribution caused by the variation of g(t) across the histogram bar,

b) a component due to the bias of $\hat{K}$ as an estimator of K, and

c) a contribution caused by the use of $\frac{1}{2}(a_i + a_{i+1})$ rather than t in the denominator of the expression for G_i.

The contributions a) and c) can, of course, be made small by reducing the bar-width, but this will increase the variance of the estimator. Little is known about the bias of $\hat{K}$ as an estimator of K; Ripley (1977) writes that 'in practice, the bias of $\hat{K}$ would seem to be small.'

ii) Choice of Parameters

We are usually interested in estimating $g(t)/\lambda^2$ for $t \in (0,a]$, where $a \leqslant \frac{1}{2} \times$ {diameter of the sampling window}; in this case, we take $a_1 = 0$ and $a_{m+1} = a$. If it is known that g(t) vanishes on an interval $(0,t_o]$ for some t_o, we may wish to make a_1 non-zero. This is the case if the pattern is hard core. If the diameter R of the discs is known, we take $a_1 = R$; otherwise, we make a_1 data-dependent and set it to be equal to the minimum inter-point distance of the pattern. This avoids having an unnecessarily low estimate in the histogram bar con-

taining R.

Having chosen the upper and lower limits a_1 and a_{m+1}, our choice of the number, m, of bars in the histogram will depend on the precision required in estimating $g(t)/\lambda^2$; expression (2.6) will be of assistance in choosing m.

The amount of smoothing applied to the histogram will also depend on the precision required, and on the visual effect desired. The choice of the smoothing parameter S is an area for further study, but some general comments can be made. In general, smoothing the histogram will increase the bias and decrease the variance of the bar estimates. For the case of points independently and uniformly distributed on a rectangular torus, considered in part i) of this section,

$$\operatorname{var}\left(G_i^{(S)}\right) = \frac{1}{2^{2S-1}N(N-1)} \left\{ \frac{1}{\pi} \sum_{j=i-S/2}^{i+S/2} \binom{S}{\frac{S}{2}+i-j} \frac{1}{a_{j+1}^2 - a_j^2} - 2^{2S} \right\} \tag{2.8}$$

for $\frac{S}{2} + 1 \leqslant i \leqslant m - \frac{S}{2}$; slightly more complicated expressions can be obtained for $1 \leqslant i \leqslant S/2$ and for $m - \frac{S}{2} - 1 \leqslant i \leqslant m$. Variance calculations for other models are complicated.

Expression (2.8) may be used to give an indication of how much a histogram should be smoothed. Another possibility is to simulate a pattern of N points independently and uniformly distributed over the sampling region D and construct smoothed histograms for this pattern. The smoothing parameter giving a satisfactory histogram for the simulated pattern can then be used when smoothing the histogram obtained from the data. A disadvantage of this approach is that, for a Poisson process, $g(t)/\lambda^2 = 1$ for all t, so that an infinite

smoothing parameter is optimal. A more satisfactory procedure is as follows: if the pattern is thought to be clustered, use an easily-simulated clustered pattern, for which $g(t)/\lambda^2$ is known - some kind of Poisson cluster process (see Chapter 0) for example - and use smoothed histograms from this simulated pattern to choose S; if the pattern is thought to be inhibited, simulate a tractable model for an inhibited process - for example, a Matérn hard core model (see section 2 of Chapter 4).

Since much of the analysis for which histograms are used is of an exploratory nature, the smoothing parameter S is often chosen by visual inspection; any model-fitting or detailed analysis is carried out using the original histogram, or a slightly smoothed version of it - see the examples in section 5 of this chapter.

iii) Asymptotic Results

Under suitable conditions, we can deduce asymptotic distributional results for our histogram estimates. The first of these is proved under sparseness conditions.

Theorem A

Let D_n, $n \in \mathbb{N}$, be a sequence of sampling windows in $\mathbb{R}^2$. For each n, let $X^{(n)}$ be a random n-point pattern in D_n, and for each n, let $\{G_{in} : 1 \leqslant i \leqslant m\}$ be the normalised second moment density histogram for $X^{(n)}$, corresponding to a partition $a_o < a_1 < \ldots < a_{m+1}$ of the interval $(a_o, a_{m+1}]$.

Suppose that

i) For each n, $X^{(n)}$ consists of n points independently and uniformly distributed over D_n.

ii) $\exists\ \lambda > 0$ s.t., for each r, $a_o \leqslant r \leqslant a_{m+1}$,

$$\lim_{n\to\infty} \binom{n}{2} \frac{m\,(Z_r(z_n))\; m(I_n)}{(m(D_n))^2} = \lambda r^2$$

for all $z_n \in I_n$, where $Z_r(z)$ denotes a disc of radius r centred at z, and $I_n = I_n(r) = D_n \setminus r$;

and

iii) for each r, $a_o \leqslant r \leqslant a_{m+1}$,

$$\lim_{n\to\infty} \binom{n}{2} \int_{B_n} \frac{m(Z_r(x) \cap D_n)}{(m(D_n))^2}\, dm = 0,$$

where $B_n = B_n(r) = D_n \setminus I_n$.

Then $\{\pi\,(a_{i+i}^2 - a_i^2)\, G_{in} : 1 \leqslant i \leqslant m\}$ converge in distribution to independent Poisson random variables having means $\lambda(a_{i+i}^2 - a_i^2)$.

Proof

Conditions ii) and iii) are the sparseness conditions of Saunders and Funk (1977). It follows from Theorem 3 of that paper that the empirical distribution function of the inter-point distances, $\{F_n(t),\ a_o < t \leqslant a_{m+1}\}$, converges weakly to a Poisson process $\{F(t) : a_o < t \leqslant a_{m+1}\}$ having mean function $E(F(t)) = \lambda t^2$. The sparseness condition iii) ensures that the expected number of pairs of points with one point lying in $B_n(a_{m+1})$ and the other within distance a_{m+1} of it, tends to 0 as n tends to infinity. It thus follows that

$$P(\hat{K}_n(t) = F_n(t),\ a_o < t \leqslant a_{m+1}) \to 1 \quad \text{as} \quad n \to \infty,$$

so that $\{\hat{K}_n(t) : a_o < t \leqslant a_{m+1}\}$ also converges weakly to a Poisson process with mean function λt^2; the result then follows. ∎

Our second asymptotic result is proved under different conditions, assuming a dense pattern.

Theorem B

Let D be a sampling window in $\mathbb{R}^2$, and let $X_1, \ldots, X_n$ be independently and unformly distributed over D. Let $\{G_i : 1 \leqslant i \leqslant m\}$ be the normalised second moment density histogram for the pattern $\{X_1, \ldots, X_n\}$, corresponding to a partition $a_o < a_1 < \ldots < a_{m+1}$ of the interval $(a_o, a_{m+1}]$. Then, as $n \to \infty$,

$$n^{\frac{1}{2}} \begin{pmatrix} G_1 - 1 \\ \vdots \\ G_m - 1 \end{pmatrix}$$ is asymptotically

normally distributed with zero mean vector.

Proof

Silverman (1976) has shown that, for n points independently and uniformly distributed over D, $\sqrt{n}\ (\hat{K}(t) - \pi t^2)$ converges weakly to a Gaussian process. The result now follows from the continuity of G_i as a function of $\hat{K}(a_{i+i})$ and $\hat{K}(a_i)$.

When we consider a torus rather than a rectangle, thus avoiding edge effects, we obtain a different result. If we assume that $X_1, \ldots, X_n$ are independently and uniformly distributed on a torus of side 2 (for convenience), then it follows from the results of Silverman (1978b) that the process $n(\sqrt{\hat{K}}(t) - \sqrt{K}(t))$ converges weakly as n tends to infinity to a non-Gaussian process with mean zero. For $t < 1$, this process has standard deviation proportional to $\sqrt{1 - \pi t^2}$. Using heuristic arguments similar to those given in Section 3 of Silverman (1978b), a Fourier expansion of the weak limit of $n(\hat{K}(t) - K(t))$

can be derived. Thus the asymptotic characteristic function of $n(\hat{K}(t) - K(t))$ can be found so that, in principle, at least, the asymptotic distribution of $\{G_i : 1 < i < m\}$ can be obtained.

4. Examples

In this section we illustrate the techniques described in section 2 on some real and simulated patterns. The examples illustrate both the applications of histogram estimates noted in section 1.

i) Lansing Woods black oaks

This dataset was described in section 6 of Chapter 1. Figure 3 shows a histogram estimate of $g(t)/\lambda^2$ for the pattern, using a bar-width of $(100\sqrt{2})^{-1} \simeq 0.00707$; the smoothed histogram with smoothing parameter S = 18 is shown in Figure 4. The histogram shows definite clustering at short ranges, with little or no interaction at longer ranges. This suggests that some kind of a Poisson cluster model might fit the pattern; such a model is fitted in the next section. Histograms using bars of constant t^2-width are shown in Figure 5; a t^2-bar-width of 0.005 was used. These graphs show the same features as those of Figures 3 and 4, suggesting clustering at short ranges, but the nature of the short-range behaviour is difficult to assess as there are few bars for small values of t.

ii) Lansing Woods maples

This dataset was also described in section 6 of Chapter 1. The histogram with bar-width $(100\sqrt{2})^{-1}$ is shown in Figure 6, and the smoothed histogram with smoothing parameter S = 18 is plotted in Figure 7. The histograms exhibit similar features to those in the previous example; again, a Poisson cluster process is fitted to the pattern in the next section.

Histograms using bars of constant t^2-width are shown in Figure 8; again, a t^2-bar-width of 0.005 was used.

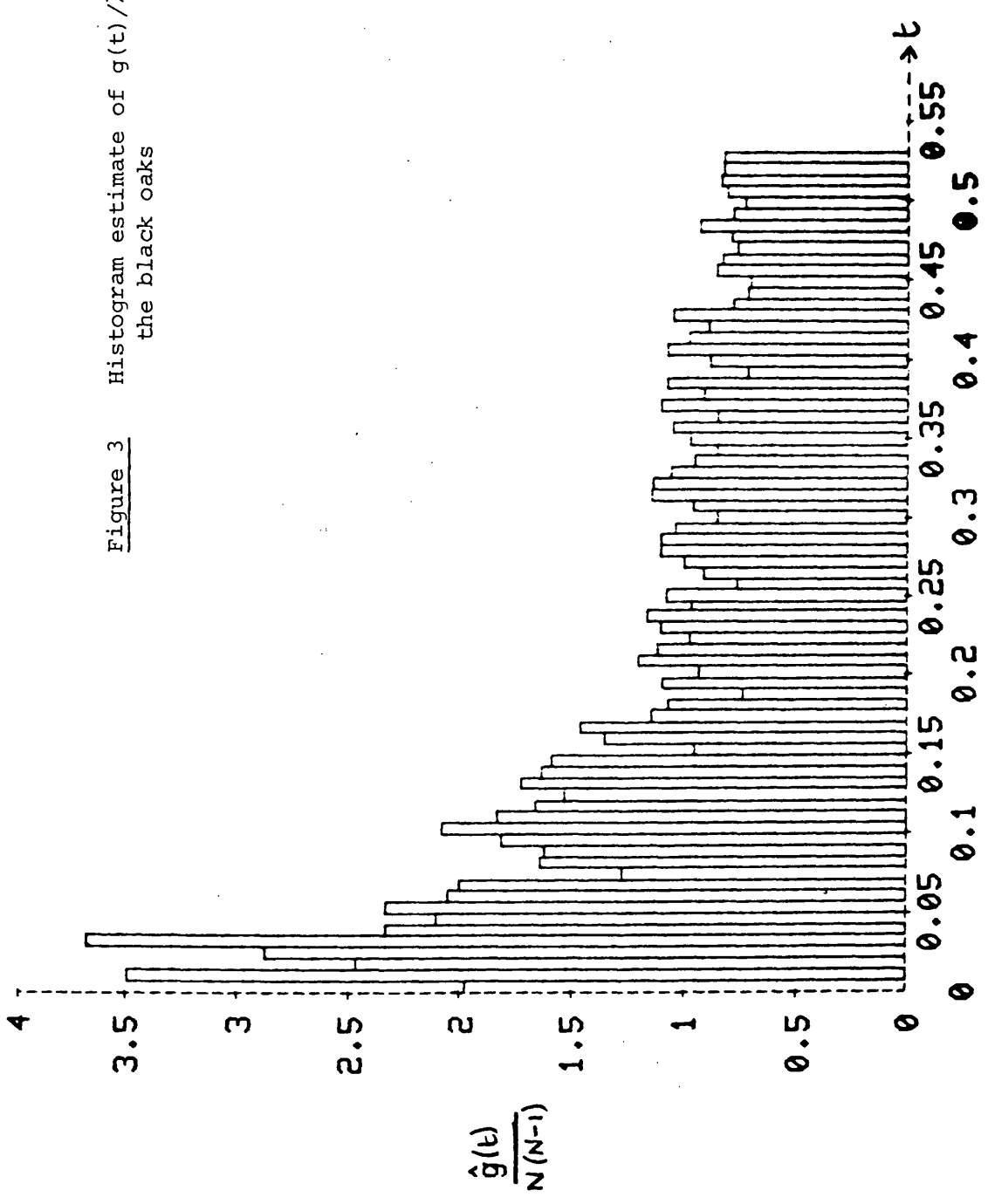


Figure 3 Histogram estimate of $g(t)/\lambda^2$ for the black oaks

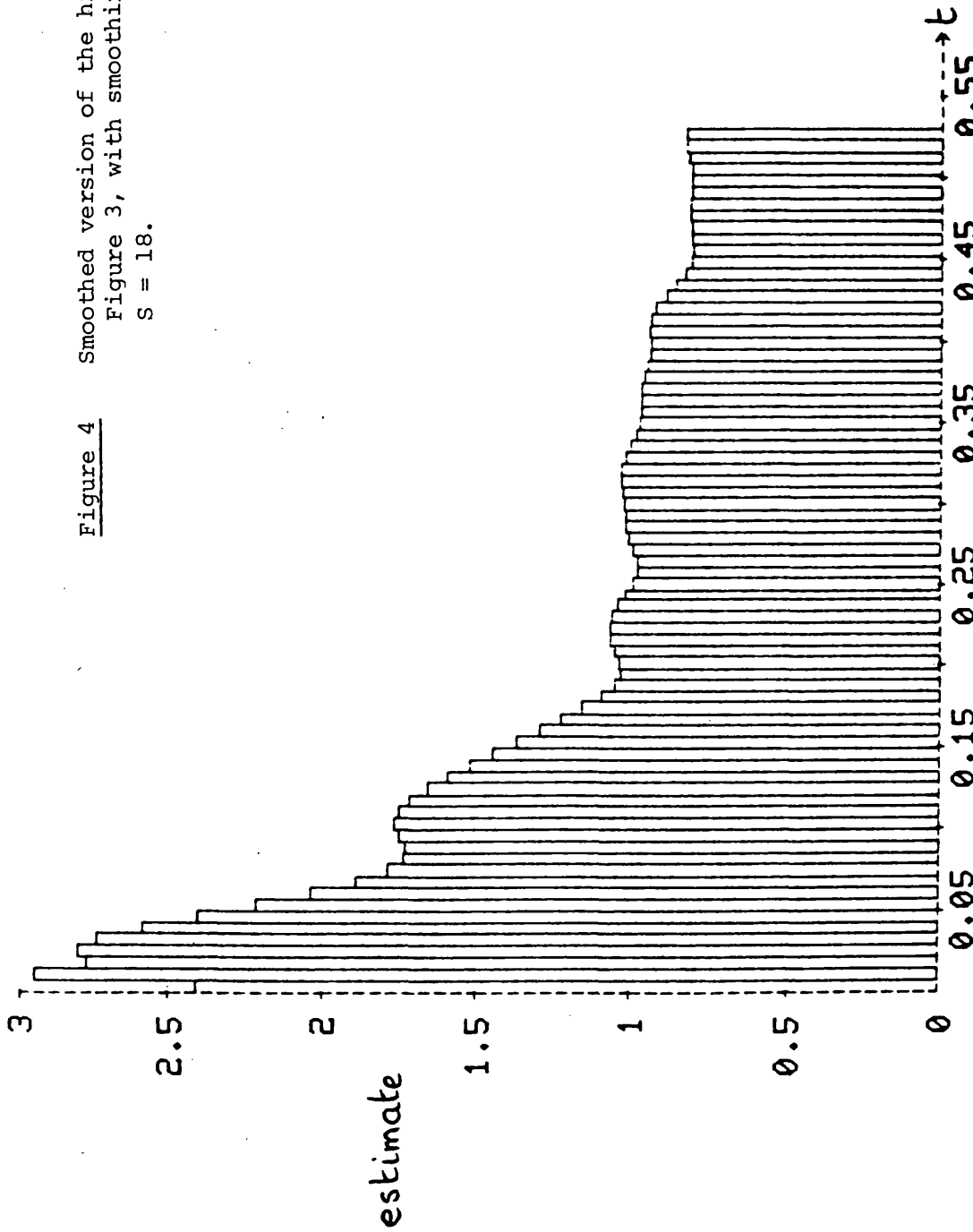
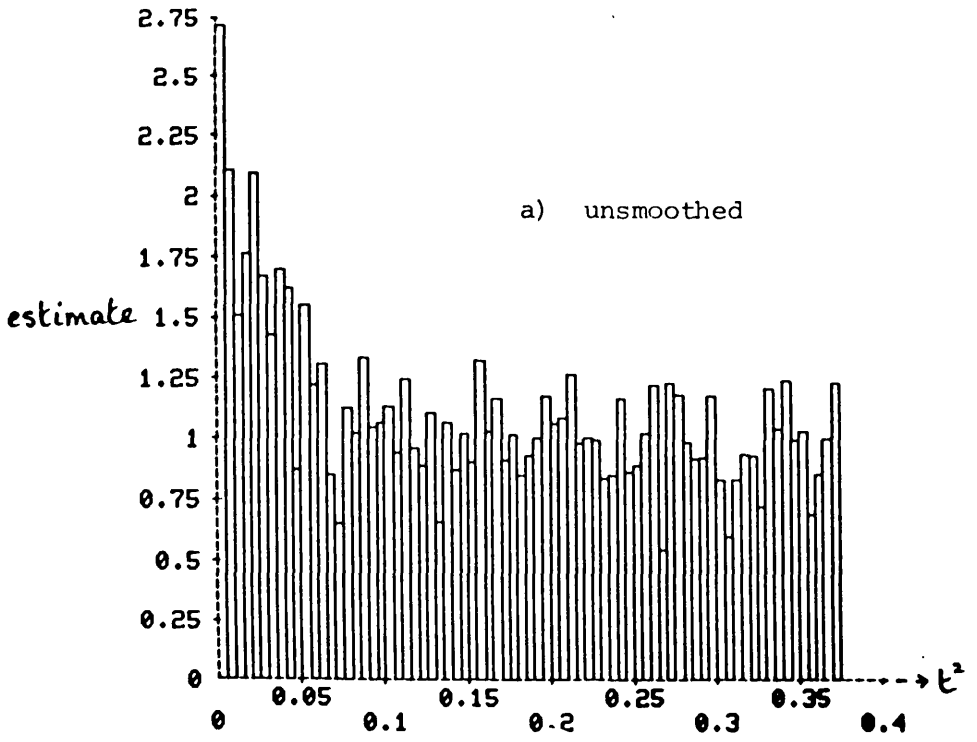
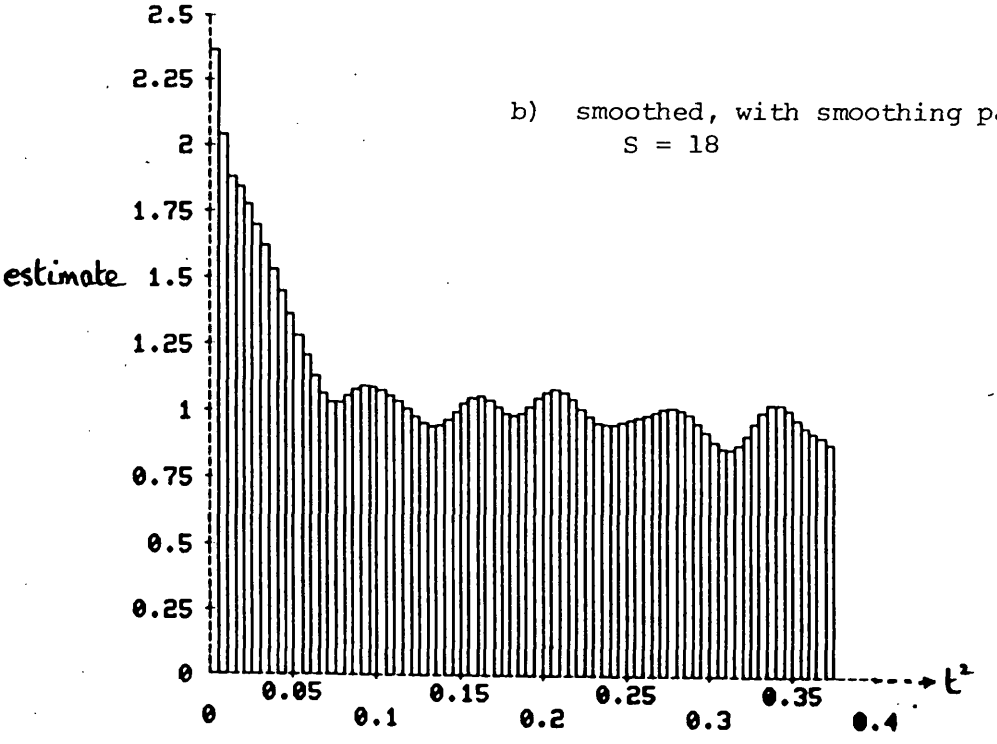


Figure 4 Smoothed version of the histogram of Figure 3, with smoothing parameter s = 18.

Figure 5 Histogram estimates with constant t^2-width for the black oaks



a) unsmoothed



b) smoothed, with smoothing parameter S = 18

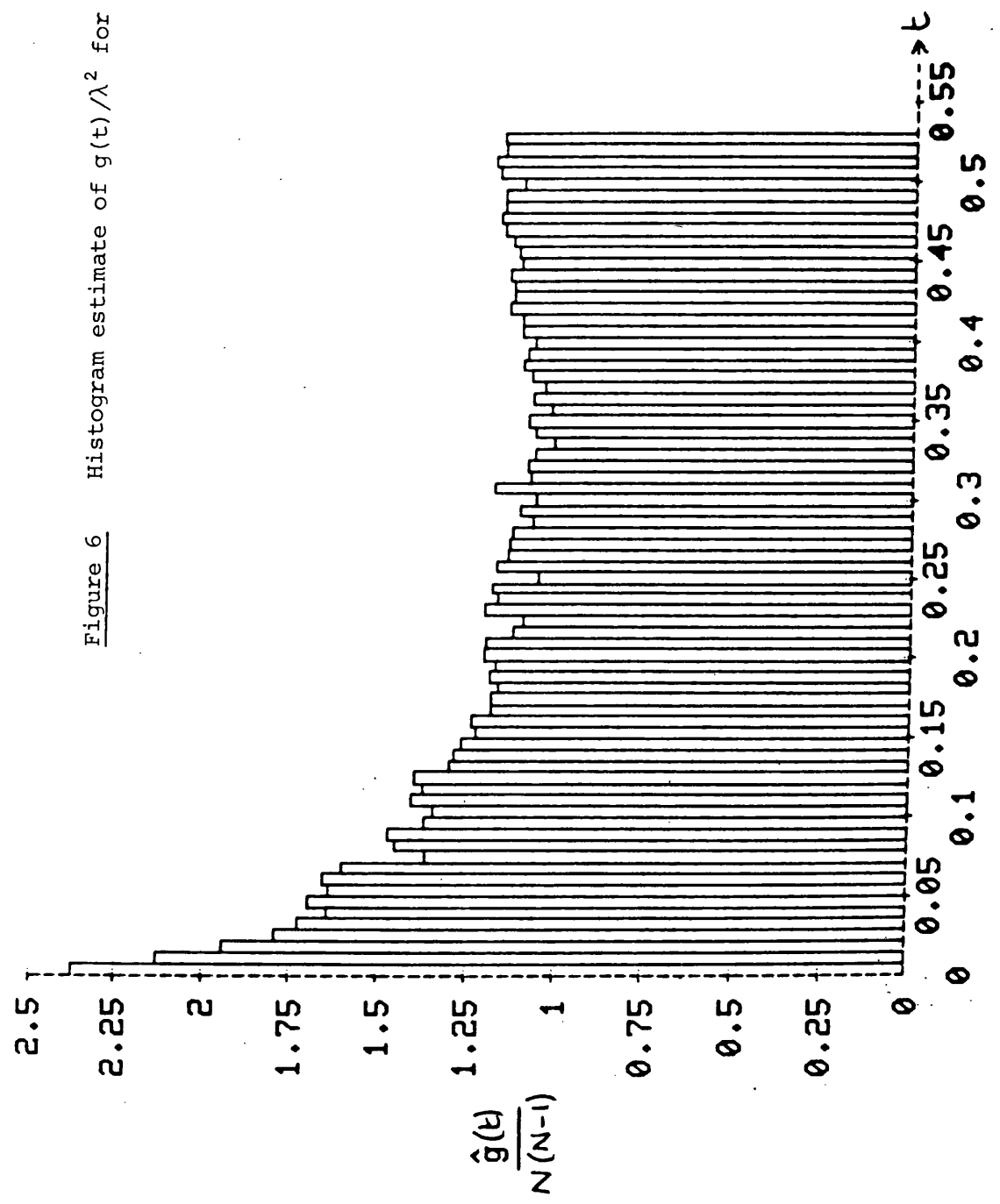


Figure 6 Histogram estimate of $g(t)/\lambda^2$ for the maples

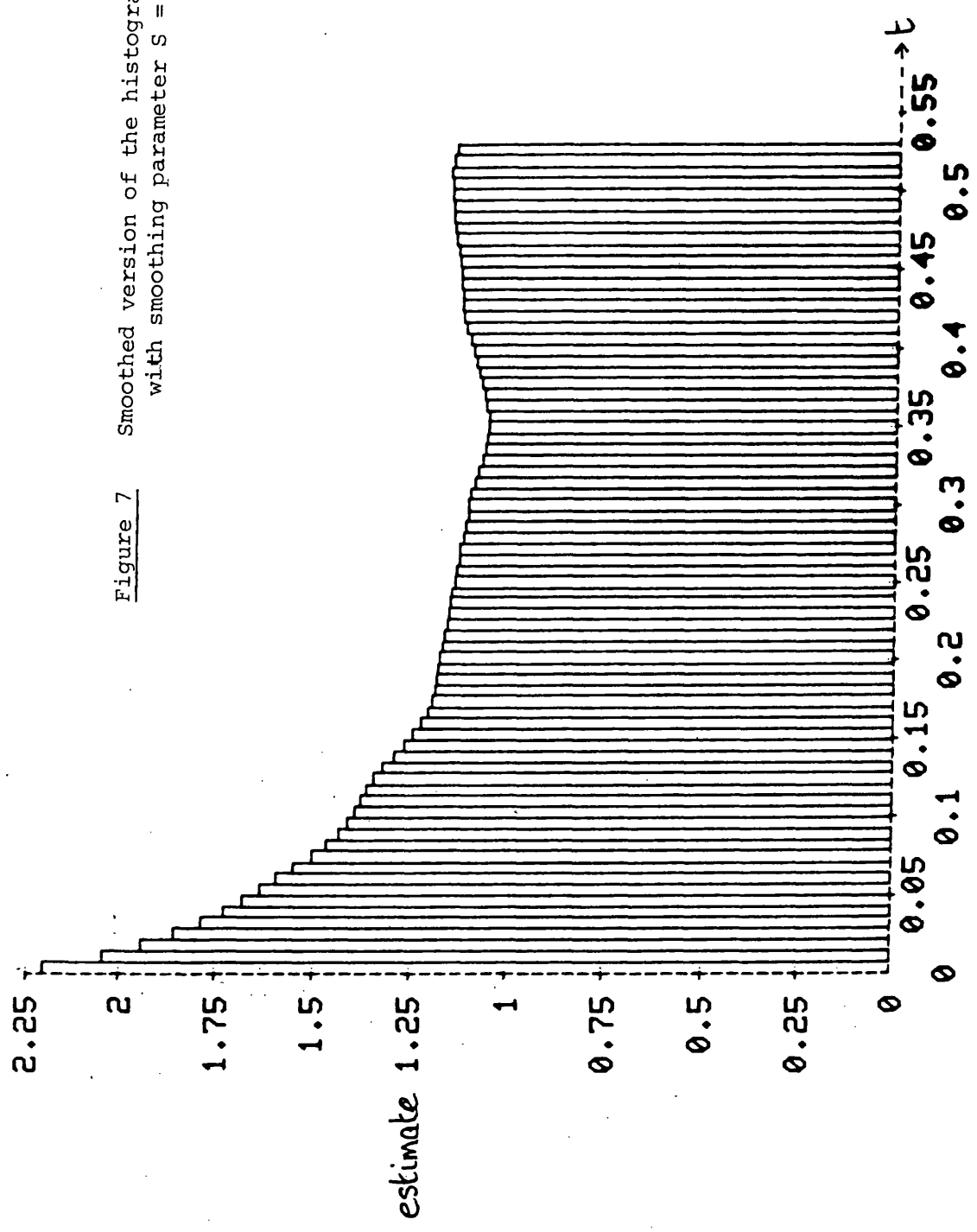
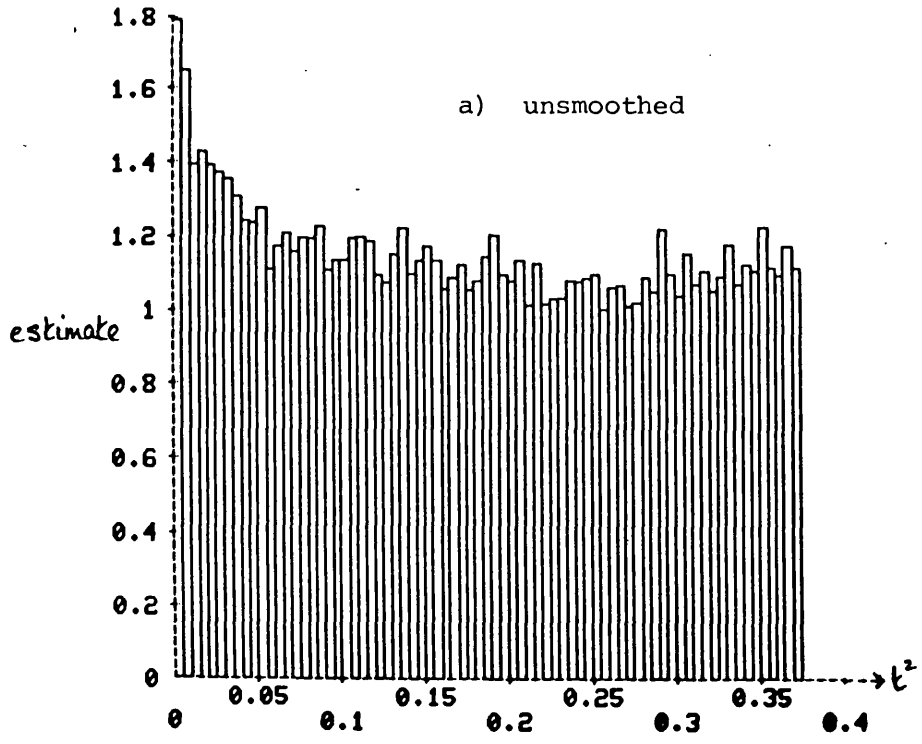
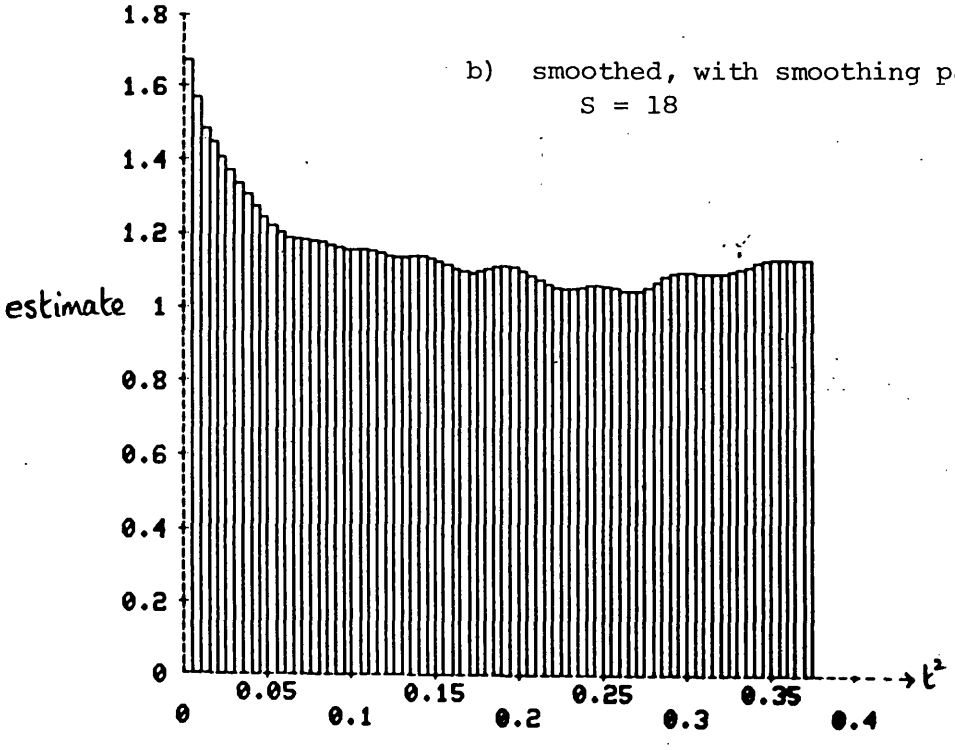


Figure 7 Smoothed version of the histogram of Figure 6, with smoothing parameter S = 18

Figure 8 Histogram estimates with constant t^2-width for the maples





iii) Histogram estimate for a Kelly-Ripley hard core model

This example illustrates the second application of histogram methods, namely the estimation of $g(t)/\lambda^2$ for known, but intractable, models. Here, we produce a histogram estimate for the Kelly-Ripley hard core model with 200 points in the unit square and disc diameter $R = 0.03$. (The model is described in section 2 of Chapter 4). The estimation was motivated by some work carried out by D.R.S. Talbot in the field of Mechanics of Random Media. This work required the estimation of $g(t)/\lambda^2$ for Kelly-Ripley hard core models with various packing densities.

The model was simulated using the hard core birth and death process described in section 2a) of Chapter 5. The birth and death process was simulated, using an initial configuration in which the points were situated at the vertices of a square grid. A histogram estimate was obtained from the pattern after 3000 steps; thereafter similar estimates were calculated at intervals of 75 steps, until 15 000 steps had been simulated in all. The histogram shown in Figure 9 is the mean of all the estimates obtained in this way.

The histogram of this example is one of many computed for Kelly-Ripley models, for various packing densities. The results obtained are expected to be of use in studying the diffusion of point defects to distributions of aligned circular cylindrical sinks in irradiated material (see Talbot and Willis (1980) for a discussion of a similar problem in three dimensions).

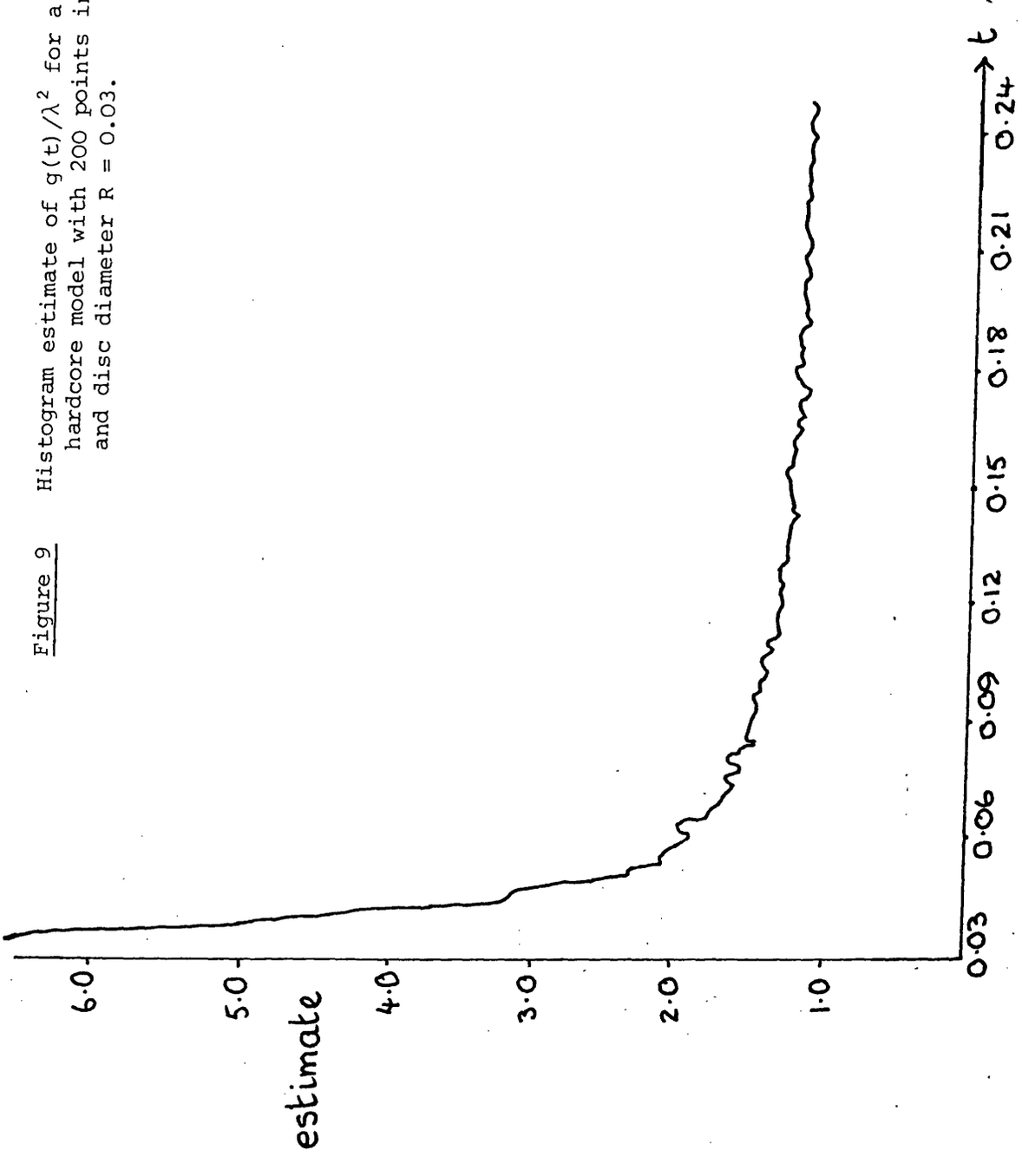


Figure 9 Histogram estimate of $g(t)/\lambda^2$ for a Kelly-Ripley hardcore model with 200 points in the unit square and disc diameter R = 0.03.

5. Fitting Poisson Cluster Processes from the Density Histogram

It was noted in section 4 that the shape of the density histograms for the Lansing Woods black oaks and maples suggested that the patterns might approximate to realisations of Poisson cluster processes. In this section, an ad hoc method is suggested for estimating the parameters of a Poisson-Poisson cluster process from the density histogram; this technique is then used to fit cluster processes to these two patterns.

We will estimate the overall intensity of the process by the sample intensity of the pattern. This leaves two quantities to be estimated: the cluster radius, R, and the number, I, of clusters present in the pattern. Having estimated these quantities, we will be able to use the method of Diggle (1979) (described in Chapter 0) for simulating Poisson cluster processes in order to assess the adequacy of the fitted model.

Figure 10 shows $g(t)/\lambda^2$ for a Poisson-Poisson cluster process with parameters (20, 10, 0.1); note that $g(t)/\lambda^2 = 1$ for $t \geq 2R$. Histograms for a simulated Poisson-Poisson cluster process with parameters R = .1, I = 20 and 200 points are shown in Figure 11. Since $t = 2R$ is the point at which the curve of $g(t)/\lambda^2$ 'flattens out' at the constant value 1, one possible estimator of R would be

$$\tilde{R} = \tfrac{1}{2} \times \left(\min \{n: G_{n+1} > G_n\} - \tfrac{1}{2}\right) \times b$$

where b is the bar-width of the histogram. Empirical evidence suggests that better results are obtained using a slightly smoothed histogram; we also want to avoid spurious troughs in the clustered part of the histogram, so we use the following modified estimator:

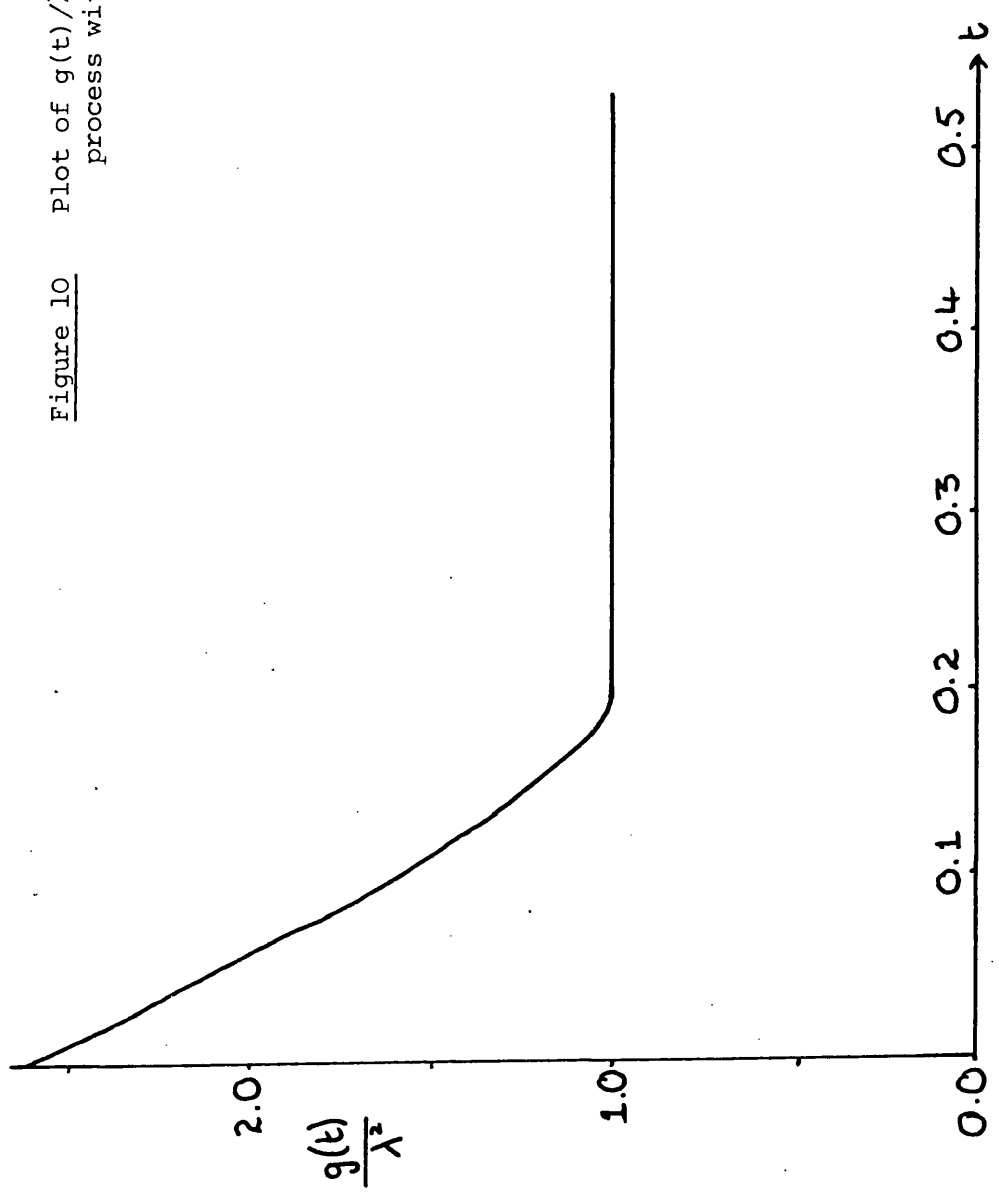
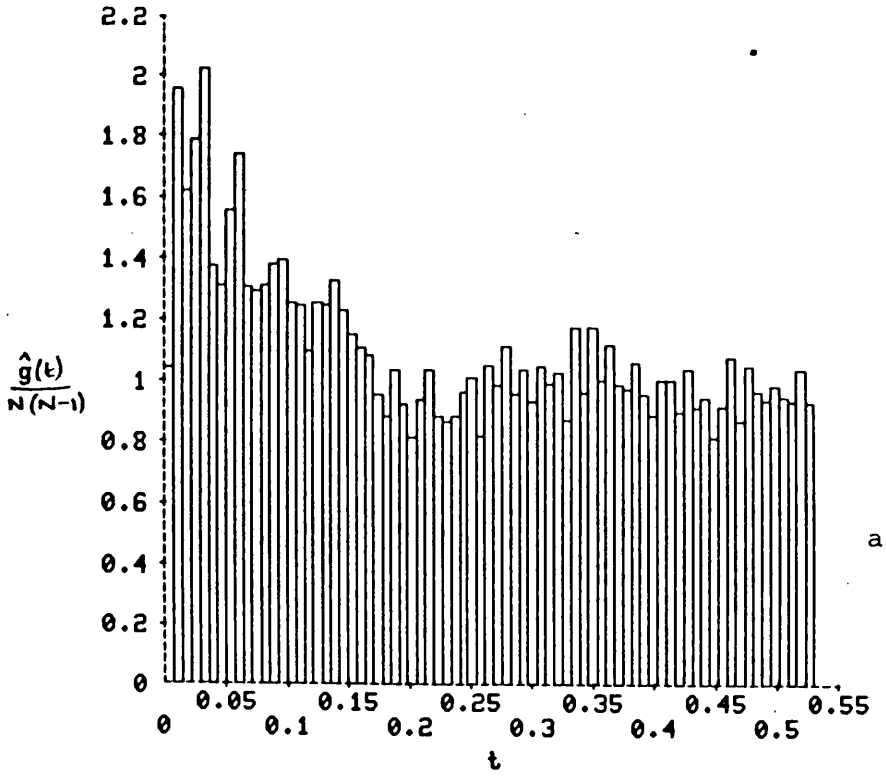
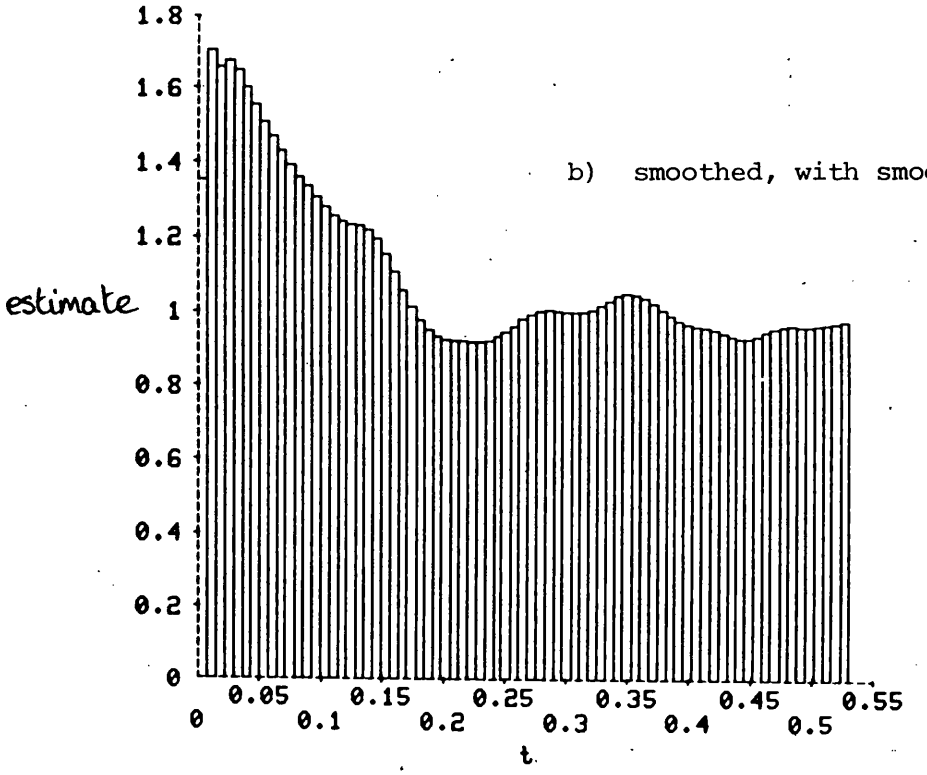


Figure 10 Plot of $g(t)/\lambda^2$ for a Poisson-Poisson cluster process with parameters (20, 10, 0.1).

Figure 11 Histogram estimates for a realisation of a 200-point Poisson-Poisson cluster process with 20 parents and cluster radius 0.1, in the unit square.



a) unsmoothed



b) smoothed, with smoothing parameter S = 18

$$\hat{R} = \tfrac{1}{2} \times (\min \{n: G^{(2)}_{n+1} > \quad G^{(2)}_{n} \text{ and } G^{(2)}_{n} < 1.2\} - \tfrac{1}{2}) \times b \qquad (2.9)$$

Our estimate of I is based on the histogram values for $t < 2\hat{R}$; I is fitted by weighted least squares. We minimise the quantity

$$S = \sum_{i=1}^{m_o} \frac{1}{w_i^2} (f(i,R,I) - G_i)^2 \; ; \qquad (2.10)$$

here, m_o is the number of bar centres less than orequal to $2\hat{R}$, and $f(i,R,I)$ is the expected value of G_i for a Poisson-Poisson cluster model with I clusters of radius R, having the same total number of points as the data. These expected values are, for simplicity, calculated assuming that the pattern is observed on a torus rather than a rectangle. The weight w_i^2 is the variance of G_i under the assumption that the points are independently and uniformly distributed on a torus; this is an approximation, but it should produce better results than those obtained using simple unweighted least squares methods. Minimisation of S over I yields the following estimator:

$$\hat{I} = \frac{\sum_{i=1}^{m_o} (\ell_i - 1)^2 / w_i^2}{\sum_{i=1}^{m_o} (1 - G_i)(\ell_i - 1)/w_i^2} \; , \qquad (2.11)$$

where $$\ell_i = \frac{2}{\pi 2\hat{R}^2} \left\{ \cos^{-1}\left(\frac{t_i}{2\hat{R}}\right) - \frac{t_i}{2\hat{R}} \sqrt{1 - t_i^2/4\hat{R}^2} \right\}$$

and $t_i = (i - \tfrac{1}{2}) \times b.$

Since we require an integer number of clusters, we use $[\hat{I}]$ or $[\hat{I}] + 1$ as our estimator of the number of clusters.

The fitting technique described above is a rough and ready one, containing many approximations, but it does provide a simple method for fitting Poisson-Poisson cluster processes. It can, of course, easily be generalised to fit other kinds of tractable Poisson cluster processes. The technique is illustrated by two examples below.

Examples

i) Lansing Woods black oaks

The histogram for this dataset (described in section 6 of Chapter 1) is shown in Figure 3. Application of the estimation procedure above yields $\hat{R} = 0.0937$ and $\hat{I} = 15.65$. The goodness of fit of such a model was assessed using empty space; 19 realisations of Poisson-Poisson cluster processes with 135 points and 15 clusters, each of radius 0.0937, were simulated in the same way as that described by Diggle (1979), and empty space statistics were calculated for each pattern. The results are shown in Figure 12; the empty space curve for the data lies inside the envelope for the simulations, and the model fits well.

ii) Lansing Woods maples

This dataset was also described in section 6 of Chapter 1; the histogram for the pattern is shown in Figure 6. Estimation of parameters gives $\hat{R} = 0.0937$ and $\hat{I} = 31.83$, suggesting 31 or 32 clusters. Figure 13 shows the empty space envelope obtained from 19 simulations using 32 clusters each of radius 0.0937. Again, a reasonable fit is achieved using this model. It is interesting to note that the two species of tree have the same fitted cluster radius of 0.0937 (about 87 feet in the original units).

These two examples illustrate the use of the fitting procedure.

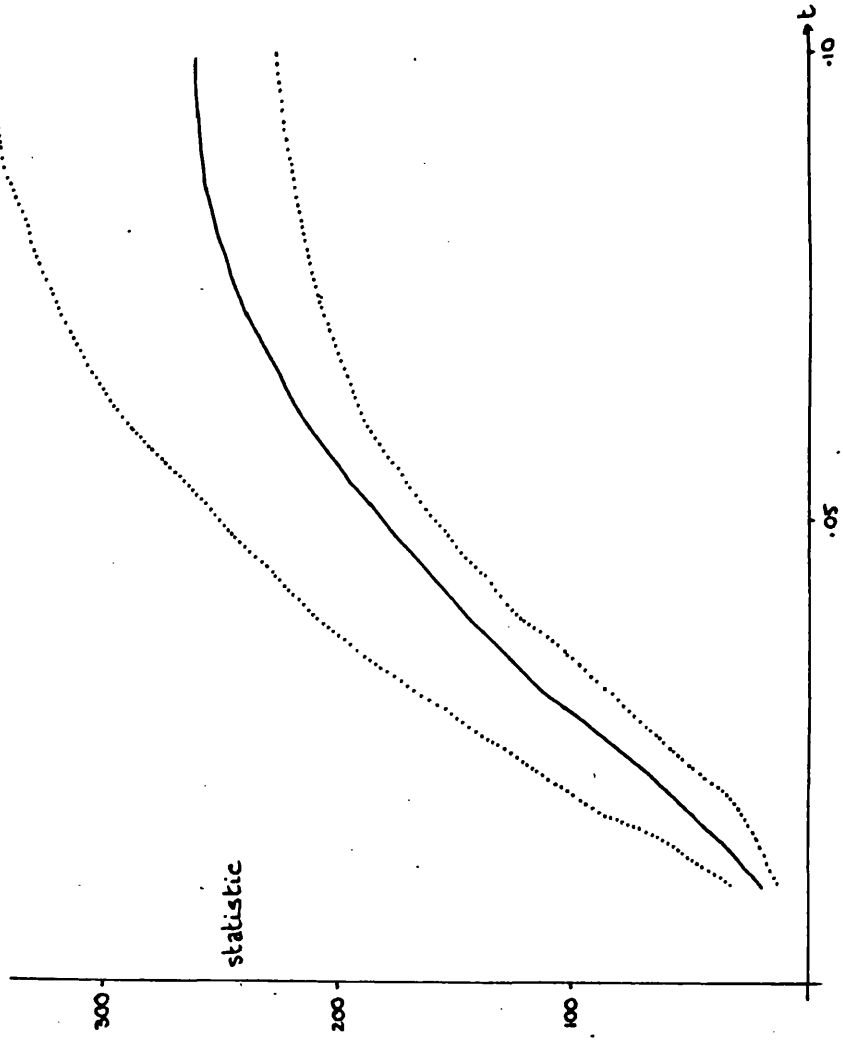


Figure 12 Fitting a cluster process to the black oaks. The statistic plotted is $\{\log (\hat{G}(U(t))) + 135\pi t^2\}/t^2$; the dotted lines are the envelope of 19 simulations of a Poisson-Poisson cluster process with 135 points, 15 parents, and cluster radius 0.0937.

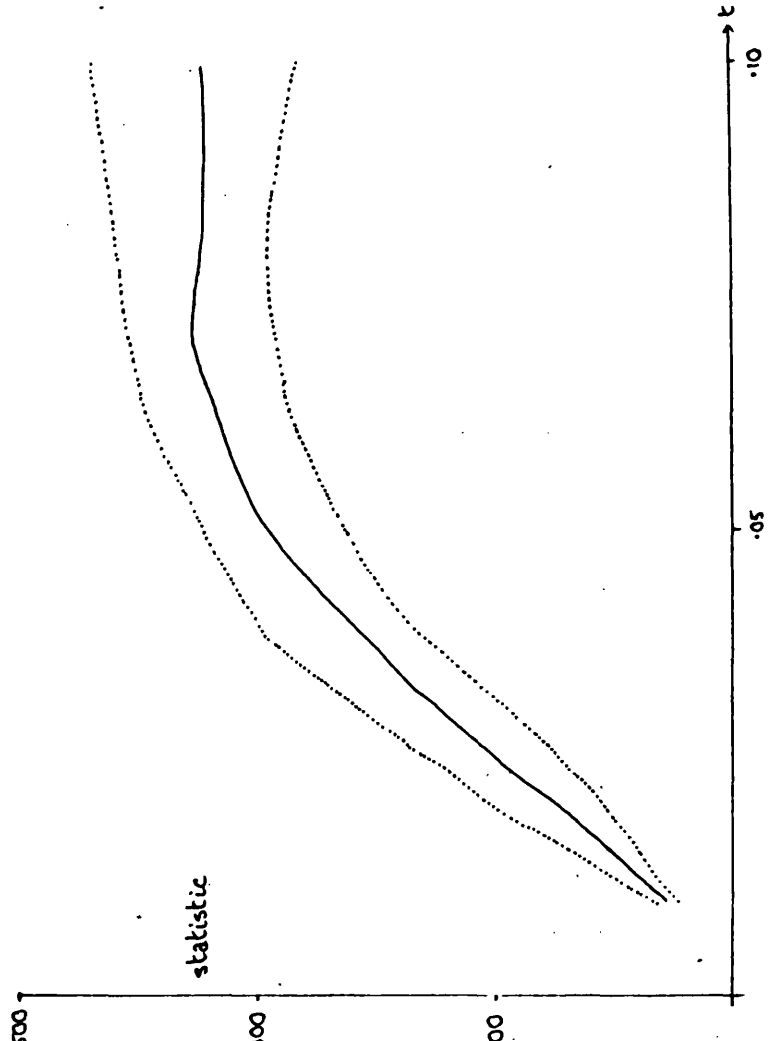


Figure 13 Fitting a cluster process to the maples. The statistic plotted is $\{\log(G(U(t))) + 514\pi t^2\}/t^2$; the dotted lines are the envelope of 19 simulations of a Poisson-Poisson cluster process with 514 points, 32 parents, and cluster radius 0.0937.

The method is an ad hoc one, but in each of the two examples, a good fit was obtained using the fitted parameter values. The results from these data suggest that the estimation procedure of this section might provide a useful method of fitting Poisson cluster processes. A fuller study of the properties of the method would require a simulation study; this is an area where further work could be carried out. Another possibility for further work is an investigation of whether the use of weighted least squares when estimating I, as described above, gives better results than ordinary least squares.

The examples also demonstrate the use of empty space statistics for assessing the goodness of fit of a model obtained using second order methods. While the same pattern is used both to fit a model and to see how well it fits, it is useful to be able to use different types of statistics at the two stages of the model-fitting - estimation of parameters and assessment of the goodness of fit of the model.

6. Discussion

The histogram estimates of the normalised second moment density described above are of use mainly at the modelling stage of the analysis of spatial patterns. While, as Ripley (1977) suggests, tests based on smoothed estimates of $g(t)$ may be less powerful than those using $\hat{K}(t)$, the first two examples of section 4 show that the use of histograms can give considerable insight into the second order properties of point patterns. In the case of these two examples, this led to the fitting of specific models to the data, using the histogram estimates.

Rather than use a histogram to estimate the normalised second moment density, we could use a smooth curve as our estimate, as is done in probability density estimation. (See, for example, Fryer (1977)). However, since we are looking mainly for pictorial representations of the second-order properties of spatial patterns, histograms provide adequate curves, and are, in some ways, easier to deal with than smooth curve estimates.

CHAPTER THREE

ANALYSING MULTITYPE PATTERNS

1. Introduction

Many problems occurring in practice involve the analysis of processes of several types of points, particularly in the plane. In this chapter, we discuss the adaptation of the empty space methods described in Chapter 1 to the analysis of multitype patterns.

There has been a considerable literature on multivariate or multitype processes; see for example several papers in Lewis (1972). Much, though by no means all, of the previous work is concerned with processes on the line or with abstract modelling considerations. Silverman, in Silverman and Lotwick (1981) has extended to multitype processes second order methods for dealing with spatial processes of points of a single type. Brown, Gereau and Jacobsen (1981) deal with a similar problem, but take a different approach. Hanisch and Stoyan (1979) have also briefly considered the second order analysis of multitype spatial processes.

In this chapter, an empty space statistic for the analysis of multitype patterns is described, and its asymptotic distribution under certain conditions is obtained. This statistic can be used for hypothesis testing and model fitting for multitype processes. The empty space methods are illustrated using practical examples in section 5, where they are compared with the second order methods of Silverman (Silverman and Lotwick, (1981)).

2. Empty space for multitype patterns

In section 2 of Silverman and Lotwick (1981), Silverman generalises to multitype point patterns some of the second order methods used for single-type patterns, and develops a method of approximately unbiased estimation of the second moment distribution $K_{xy}(t)$. Here, we describe a complementary method of analysing multitype patterns, based on the empty space methods described in Chapter 1.

Let X and Y be two stationary point processes in the plane; denote the combined process by XUY. If the two processes X and Y are independent, we have the following identity holding for all Borel subsets A of the plane:

$$G_{XUY}(A) = G_X(A) \times G_Y(A) \tag{3.1}$$

This can be more conveniently expressed in the form

$$\log (G_{XUY}(A)) = \log (G_X(A)) + \log (G_Y(A)). \tag{3.2}$$

Consequently, if we have observed a process of x-points and y-points, and we wish to test for the independence of the X- and Y- processes, or examine the interaction between the two types, we use as our statistic

$$T_{xy}(A) = \log (\tilde{G}_{xUy}(A)) - \log (\tilde{G}_x(A)) - \log (\tilde{G}_y(A))$$

or, more usually, a multivariate statistic $(T_{xy}(A_1), \ldots, T_{xy}(A_n))$, where the A_i's are discs of different radii. (For square or rectangular test sets we would, of course, use $\hat{G}$ rather than $\tilde{G}$).

To give an overall picture of the interaction between the two processes, we plot $T_{xy}(U(t))$ against t, in a way analogous to that used for single-type empty space. Positive values of T_{xy} would indicate attraction between the x- and y- points, whereas negative values would indicate inhibition. Some distributional results, which enable us to use T_{xy} to test for independence of the processes in certain circumstances, are given in section 3 below. The use of T_{xy} is illustrated with examples in section 5.

Attention need not be restricted to consideration of the types two at a time; statistics analogous to T_{xy} can be used to investigate higher-order 'interactions' between types of points. More details of this are given in section 6.

3. Distributional Results

As for the single-type case, asymptotic distributional results can be proved for multitype empty space statistics, under certain conditions, as the sampling window D increases to $\mathbb{R}^2$.

Theorem A

Let A be a compact measurable subset of $\mathbb{R}^2$. Let X and Y be two stationary point processes in the plane.

Define $U_k = [0,k] \times [0,k] \subseteq \mathbb{R}^2$, and

$$F_A^{(z)}(U) = \int_{\underset{\sim}{u} \in U} I(N_z(A_{\underset{\sim}{u}}) = 0)\, dm(u) - G_z(A)m(u)$$

for Z = X,Y or X∪Y.

Define $$\beta_A(U', U'') = \sup_{\substack{A \in \mathcal{U}(U') \\ B \in \mathcal{U}(U'')}} |P(A \cap B) - P(A)P(B)|$$

where $\mathcal{U}(U)$ is the σ-algebra of events generated by the random variables $(F_A^{(X \cup Y)}(C), F_A^{(X)}(C), F_A^{(Y)}(C))$ for all measurable subsets C of the set U, and let

$$\beta_A^*(r) = \sup_{\rho(U', U'') \geq r} \beta_A(U', U'') \quad .$$

Let $$G_k^{(z)}(A) = \frac{F_A^{(z)}(U_k)}{m(U_k)} \qquad k = 1,2, \ldots$$

and let

$$T_k(A) = \log G_k^{(X \cup Y)}(A) - \log G_k^{(X)}(A) - \log G_k^{(Y)}(A).$$

Then, provided $G_{X \cup Y}(A), G_X(A), G_Y(A) > 0$ and

$$\beta_A^*(r) = O(r^{-(2+\eta)}) \quad \text{for some } \eta > 0 \,, \tag{3.3}$$

we have

$$(m(U_k))^{\frac{1}{2}} T_k(A) \xrightarrow{\mathcal{D}} N(0,\sigma^2) \quad \text{as } k \to \infty ,$$

where
$$\sigma^2 = \frac{V_{X\cup Y}}{\{G_{X\cup Y}(A)\}^2} + \frac{V_X}{(G_X(A))^2} + \frac{V_Y}{(G_Y(A))^2}$$

$$- \frac{2C_{X\cup Y,X}}{G_X(A)\ G_{X\cup Y}(A)} - \frac{2C_{X\cup Y,Y}}{G_Y(A)\ G_{X\cup Y}(A)} + \frac{2C_{X,Y}}{G_X(A)\ G_Y(A)} ,$$

where
$$V_z = \int_{\underset{\sim}{u}\in\mathbb{R}^2} \{P(N_z(A_{\underset{\sim}{0}}) + N_z(A_{\underset{\sim}{u}}) = 0) - G_z(A)^2\}\, dm(\underset{\sim}{u})$$

and

$$C_{z_1,z_2} = \int_{\underset{\sim}{u}\in\mathbb{R}^2} \{P(N_{z_1}(A_{\underset{\sim}{0}}) + N_{z_2}(A_{\underset{\sim}{u}}) = 0) - G_{z_1}(A)G_{z_2}(A)\}\, dm(\underset{\sim}{u})$$

Proof

The proof is in two parts; first we show that

$(m(U_k))^{\frac{1}{2}} \begin{pmatrix} G_k^{(X\cup Y)}(A) \\ G_k^{(X)}(A) \\ G_k^{(Y)}(A) \end{pmatrix}$ is asymptotically normally distributed, then

we apply a standard theorem to show that the transformed statistic $T_k(A)$ has the required limiting distribution.

To prove the first part, we show that for any $\underset{\sim}{a} = (a_1\ a_2\ a_3) \in \mathbb{R}^3$ with $\underset{\sim}{a} \neq \underset{\sim}{0}$,

$$(m(U_k))^{\frac{1}{2}}(a_1 G_k^{(X\cup Y)}(A) + a_2 G_k^{(X)}(A) + a_3 G_k^{(Y)}(A))$$

$$\xrightarrow{\mathcal{D}} N(0, \underset{\sim}{a}^T \, \sharp \, \underset{\sim}{a}) \quad \text{as } k \to \infty$$

where

$$\Sigma = \begin{pmatrix} V_{XUY} & C_{XUY,X} & C_{XUY,Y} \\ C_{XUY,X} & V_X & C_{X,Y} \\ C_{XUY,Y} & C_{X,Y} & V_Y \end{pmatrix}$$

The proof of this is analogous to those of Theorems A and B of Chapter 1; all the conditions necessary for the application of Lemma 1 of Chapter 1 can be seen to hold. It thus follows that, as $k \to \infty$,

$$(m(U_k))^{\frac{1}{2}} \begin{pmatrix} G_k^{(XUY)}(A) \\ G_k^{(X)}(A) \\ G_k^{(Y)}(A) \end{pmatrix} \xrightarrow{\mathcal{D}} N(O, \Sigma).$$

Application of the Lemma on page 207 of Rao (1952), with $f(t_1, t_2, t_3) = \log t_1 - \log t_2 - \log t_3$, and using k^2 rather than n, completes the proof of the theorem. ∎

Note that condition (3.3) is guaranteed to hold if each of X and Y satisfies (1.9) and the X and Y processes are independent; in this case, the expression for σ^2 reduces to

$$\frac{V_{XUY}}{(G_X(A))^2 (G_Y(A))^2} - \frac{V_X}{(G_X(A))^2} - \frac{V_Y}{(G_Y(A))^2} \qquad (3.4)$$

The theorem can also be applied in cases when the marginal processes are not independent; for example, the conditions can be seen to hold for the random Dirichlet cell processes of Brown, Milne and Silverman, (1981, section 2).

In exactly the same way as in section 4 of Chapter 1, we can extend Theorem A to deal with multivariate empty space statistics, and so obtain the following

Theorem B

Let $A_1, \ldots, A_n$ be compact measurable subsets of $\mathbb{R}^n$. Let X be a stationary point process in the plane. Define U_k, $G_k(A_i)$ and $\beta^*_{A_i}(r)$ as in Theorem A. Then, provided that, for $i = 1, \ldots, n$,

$$\beta^*_{A_i}(r) = O(r^{-(2+\eta)}) \quad \text{for some } \eta > 0 , \tag{3.5}$$

we have

$$(m(U_k))^{\frac{1}{2}} \begin{pmatrix} T_k(A_1) \\ \cdot \\ \cdot \\ \cdot \\ T_k(A_n) \end{pmatrix} \xrightarrow{\mathcal{D}} N(O, \Sigma)$$

where

$$\Sigma = (\sigma_{ij}) , \quad 1 \leqslant i,j \leqslant n$$

and

$$\begin{aligned} \sigma_{ij} = {} & \frac{W^{ij}_{X\cup Y, X\cup Y}}{G_{X\cup Y}(A_i)G_{X\cup Y}(A_j)} + \frac{W^{ij}_{X,X}}{G_X(A_i)G_X(A_j)} + \frac{W^{ij}_{Y,Y}}{G_Y(A_i)G_Y(A_j)} \\ & - \frac{W^{ij}_{X\cup Y, X}}{G_{X\cup Y}(A_i)G_X(A_j)} - \frac{W^{ij}_{X\cup Y, Y}}{G_{X\cup Y}(A_i)G_Y(A_j)} - \frac{W^{ij}_{X, X\cup Y}}{G_X(A_i)G_{X\cup Y}(A_j)} \\ & - \frac{W^{ij}_{Y, X\cup Y}}{G_Y(A_i)G_{X\cup Y}(A_j)} + \frac{W^{ij}_{X,Y}}{G_X(A_i)G_Y(A_j)} + \frac{W^{ij}_{Y,X}}{G_Y(A_i)G_X(A_j)} ; \end{aligned} \tag{3.6}$$

here, $$W^{ij}_{z_1,z_2} = \int_{\underset{\sim}{u}\in\mathbb{R}^2} \{P(N_{z_1}(A_{i\underset{\sim}{u}}) + N_{z_2}(A_{j\underset{\sim}{u}}) = O) - G_{z_1}(A_i)G_{z_2}(A_j)\}\, dm(\underset{\sim}{u}).$$

(3.7)

4. Use of the Statistics

In this section we discuss the use of empty space statistics for analysing multitype patterns, in particular for hypothesis testing and model fitting. These remarks are illustrated with some examples in section 5.

T_{xy} can be used in conjunction with $\tilde{G}_x$ and $\tilde{G}_y$, and also with second order statistics, to provide insight into the behaviour of the processes. The statistic $T_{xy}(U(t))/t^3$ can be plotted; under the double Poisson hypothesis, this statistic has, for low values of t, approximately constant variance. If the processes are independent, then $T_{xy}(U(t)) \simeq 0$. A tendency for x- and y- points to attract up to scale t is demonstrated by $T_{xy}(U(t)) > 0$, while inhibition is demonstrated by $T_{xy}(U(t)) < 0$.

A null hypothesis of particular interest is the hypothesis that the processes are independent. The distribution of T_{xy} will of course depend on the marginal structure of the X- and Y- processes; if we are willing to make certain assumptions about this marginal structure, the results of the previous section can be used to provide acceptance regions for the null hypothesis, as follows. If the marginal distributions of the X- and Y- processes are such that, if the two processes are independent, the conditions of Theorem 1 of this chapter are satisfied, then, on the null hypothesis, T_{xy} will be asymptotically Normally distributed. Thus, if we are prepared to assume that, for example, each of the X- and Y- processes is marginally a Poisson or Poisson cluster process, then we can use the asymptotic Normality of T_{xy} to provide approximate acceptance regions for the null hypothesis; these can be drawn on the graph of $T_{xy}(U(t))$. On the hypothesis of independence, the asymptotic mean of T_{xy} will be zero. The asymptotic variance will

depend on the parameters of the X- and Y- processes; if these are unknown, they will need to be fitted, and the variance calculated from the fitted values. The goodness of the fit can be assessed by looking at empty space or distance statistics for the marginal processes individually. These devices mean that the acceptance regions are only approximate, but they do provide guidelines as to the plausibility of the independence hypothesis.

If we are not willing to make assumptions about the marginal structure of the processes, we need another device for assessing the interaction between them. Such a method, the use of Monte Carlo tests conditional on the observed marginal structure, is described in section 4 of Silverman and Lotwick (1981). The details are as follows. Suppose that we are observing the processes on a square or rectangular region. By imposing periodic boundary conditions, wrap the region of interest onto the torus, so that points on opposite edges of the rectangle are considered to be near. Take the observed processes and keep the x-process fixed throughout. Generate a new y-process by translating the y-process through a random amount on the torus, the random translation being chosen from the uniform distribution on translations of the torus. The resulting y-process will have exactly the same marginal second and higher order structure as the original observed process and will be stationary and independent of the x-process. By considering y-processes translated by different independent amounts, Monte Carlo acceptance regions for any translation-invariant statistics for tests of the hypothesis of independence can be constructed. Because of the translation invariance of the statistics, the distribution obtained will be the same as that for the full stationary model obtained by randomly translating both the processes by independent amounts.

Multitype empty space statistics can be used for model building in the same way as for the single-type case, except that now, there are three statistics being used - two for the marginal processes and one for the interaction between them. Such model fitting is illustrated in the examples of the next section.

5. Examples

In this section, we illustrate some of the techniques for analysing multitype patterns; both the empty space methods described above and Silverman's second-order methods (Silverman and Lotwick (1981)) are applied to three datasets. The first two patterns considered form part of some data collected by Professor T. Takahashi of the University of Sendai, Japan (Takahashi, 1970), and the third is again taken from the Lansing Woods data of Professor D.J. Gerrard (see Gerrard (1969)).

Example 1 : Cerebral Cortex

The dataset for this example (shown in Figure 1) consists of a cross-section through part of a human cerebral cortex, with the positions of the veins and arteries marked, and can thus be regarded as a realisation of a two-type process, the two types corresponding to veins and arteries.

The interaction between veins and arteries was first investigated using confidence bands derived from the double Poisson hypothesis. Before doing this, it was necessary to check that the marginal distributions were approximately Poisson. This was done using both empty space and second-order methods; the results are shown in Figures 2 - 5. (All empty space statistics in the three examples in this section were computed using a circular test set). The positions of the veins can be accepted as a Poisson process; the arteries show a slight tendency towards inhibition, the data curves crossing the Poisson confidence bands in some places. Inhibition in the marginals would make the $\hat{K}$ confidence bands for a test of independence conservative (see section 2.4 of Silverman and Lotwick (1981)).

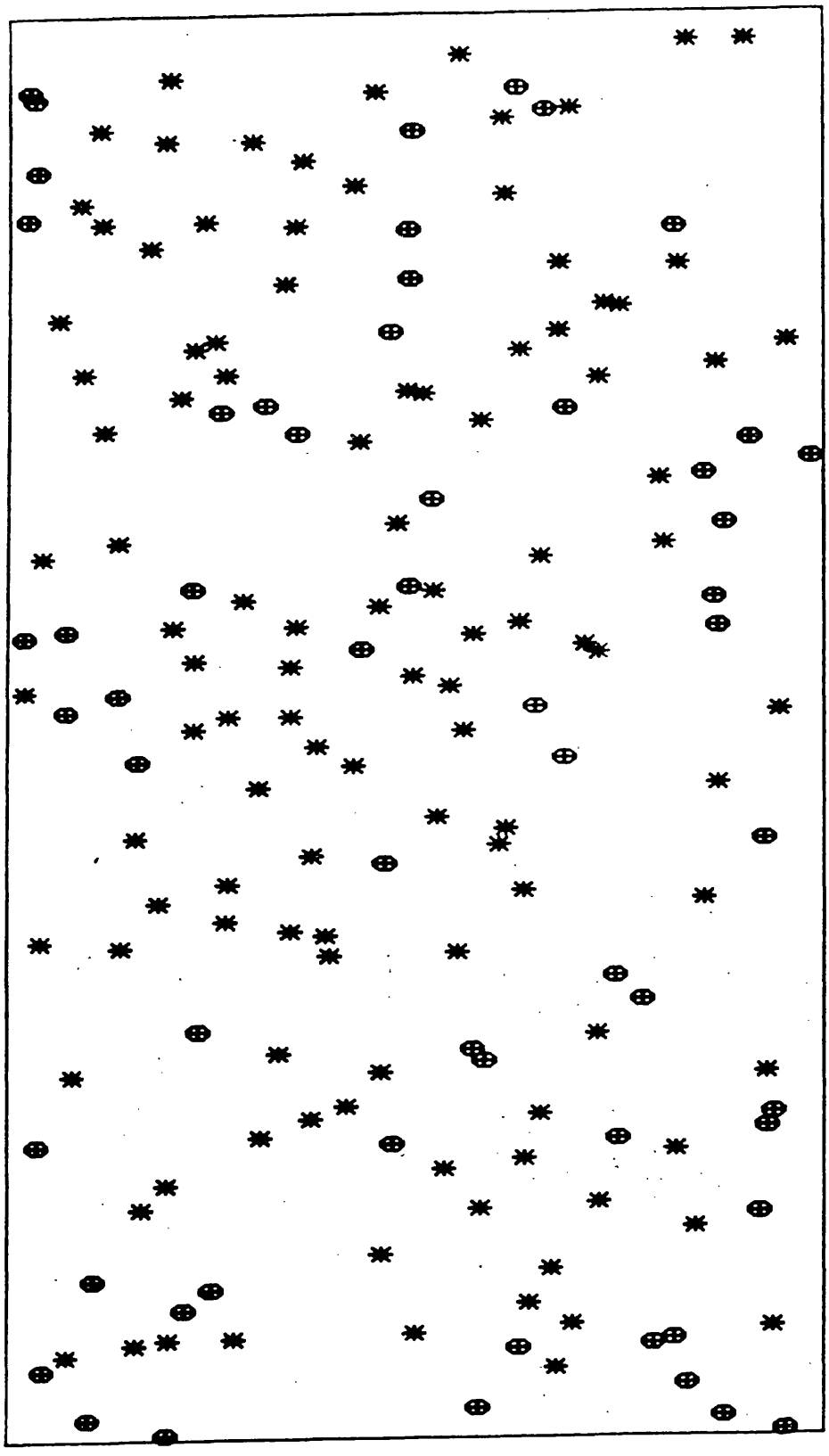


Figure 1 A 28.6 × 16.25 rectangle containing the positions in a cross-section through a human cerebral cortex of 112 arteries (*) and 58 veins (⊕).

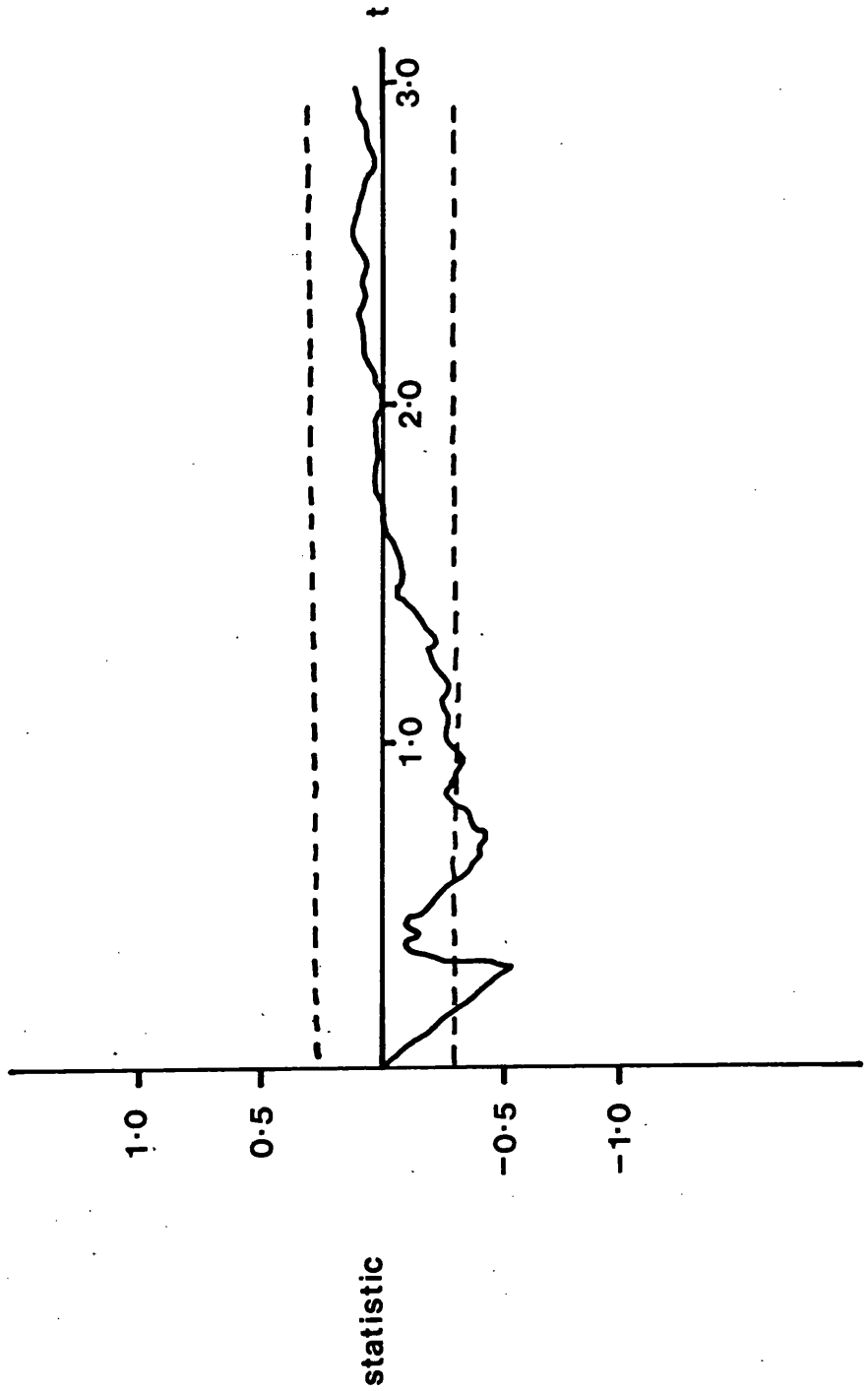


Figure 2 Cerebral cortex. $\hat{K}$ for the arteries alone. The statistic plotted is $\sqrt{\hat{K}(t)} - \sqrt{\pi}t$. The dashed curves are estimated 95% confidence bands computed for a Poisson process with corresponding intensity.

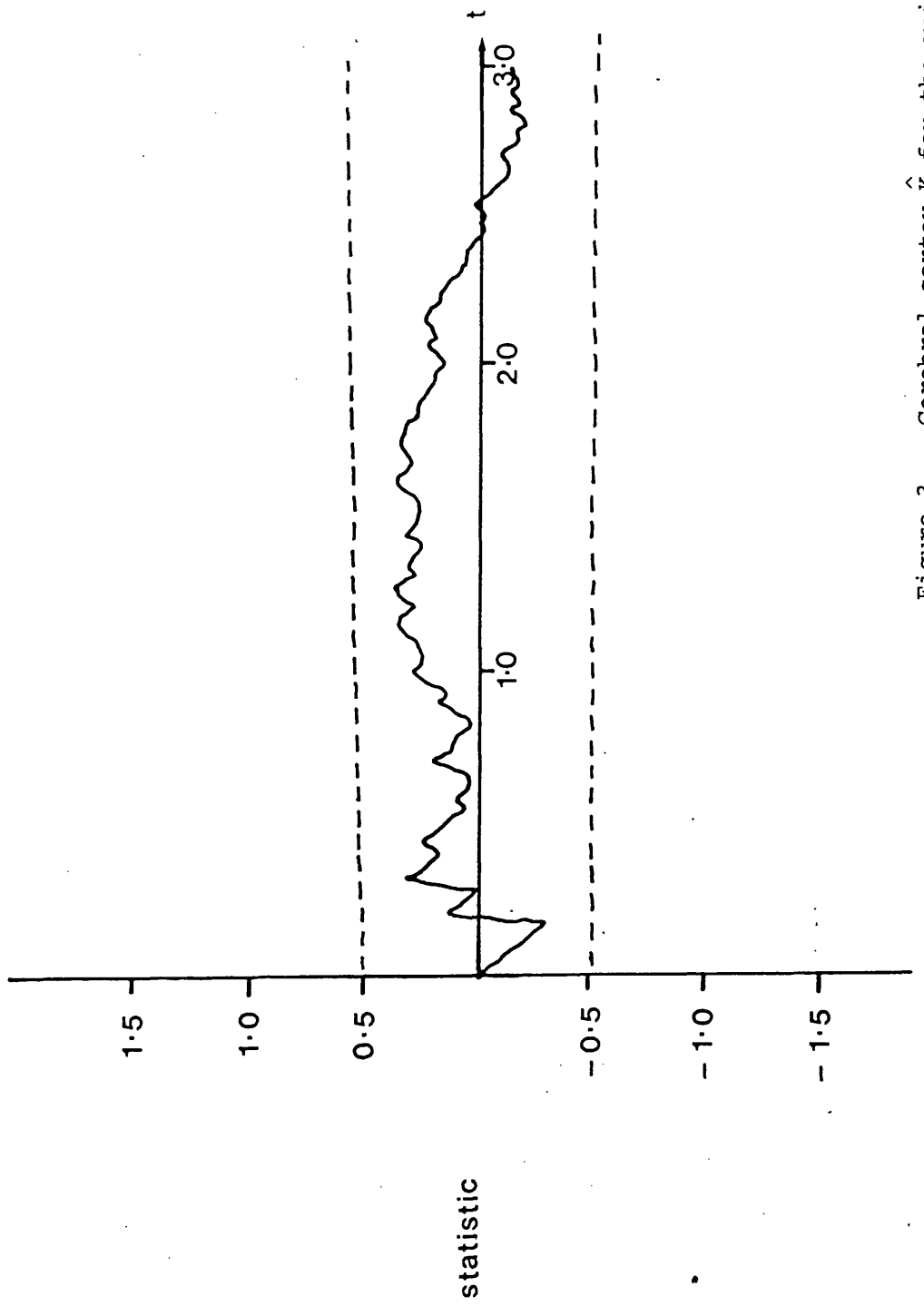


Figure 3 Cerebral cortex. $\hat{K}$ for the veins alone, with estimated 95% confidence bands for a Poisson process. The statistic plotted is $\sqrt{\hat{K}(t)} - \sqrt{\pi t}$.

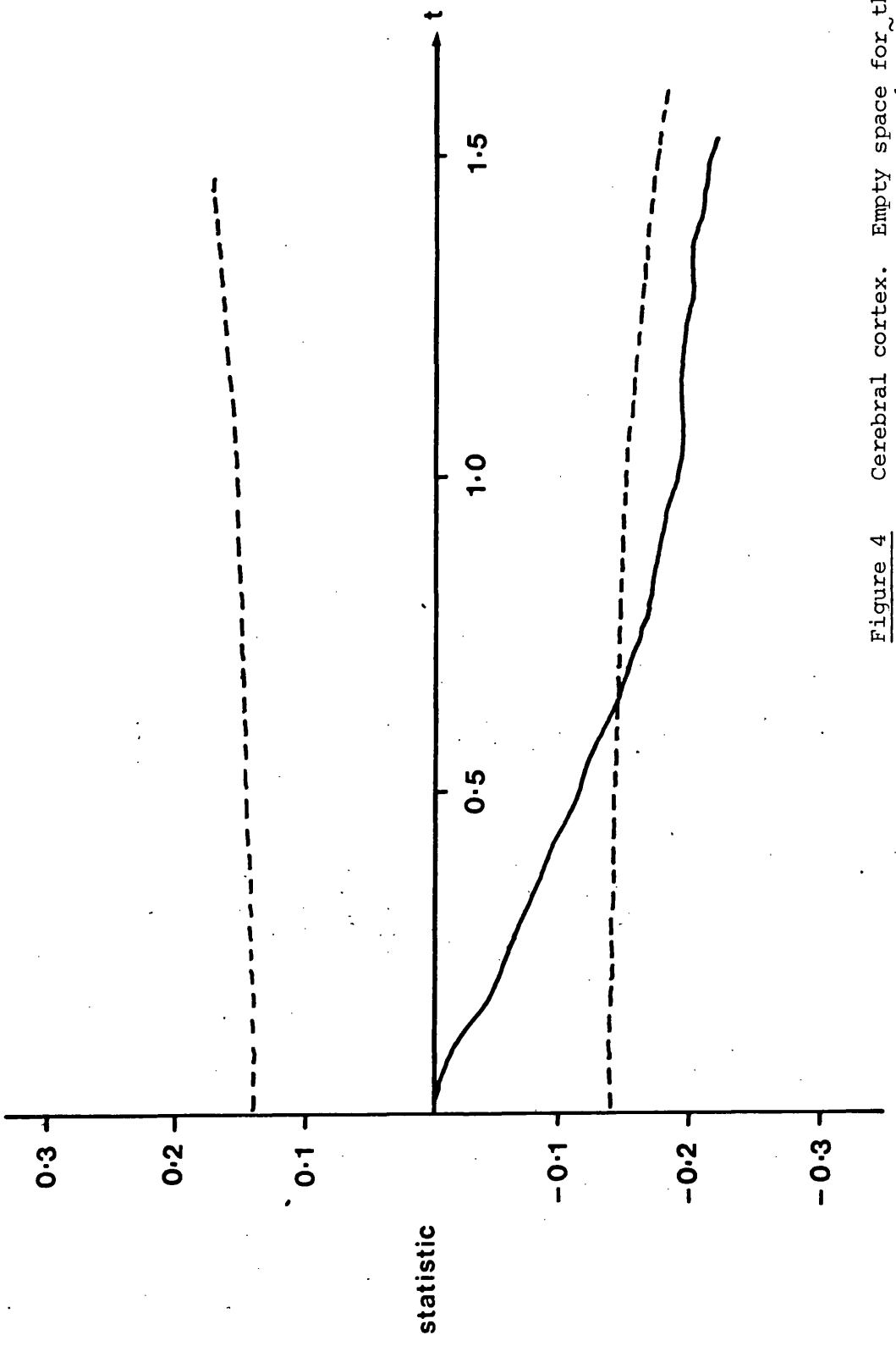


Figure 4 Cerebral cortex. Empty space for the arteries alone. The statistic plotted is $\{\log(G(U(t)) + .241\pi t^2)\}/t^2$. The dashed lines are asymptotic 95% confidence bands computed for a Poisson process.

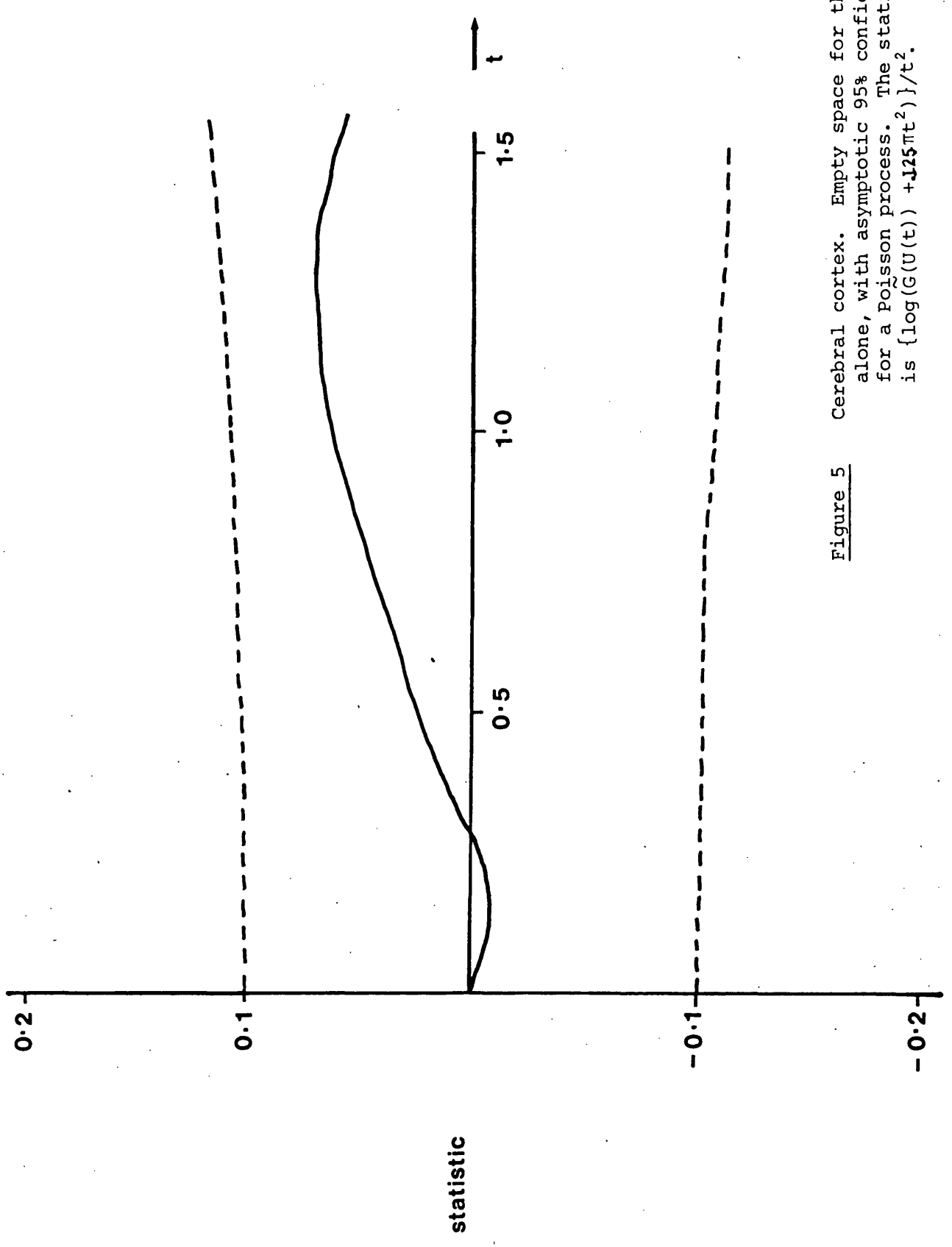


Figure 5 Cerebral cortex. Empty space for the veins alone, with asymptotic 95% confidence bands for a Poisson process. The statistic plotted is $\{\log(\tilde{G}(U(t)) + 1.25\pi t^2)\}/t^2$.

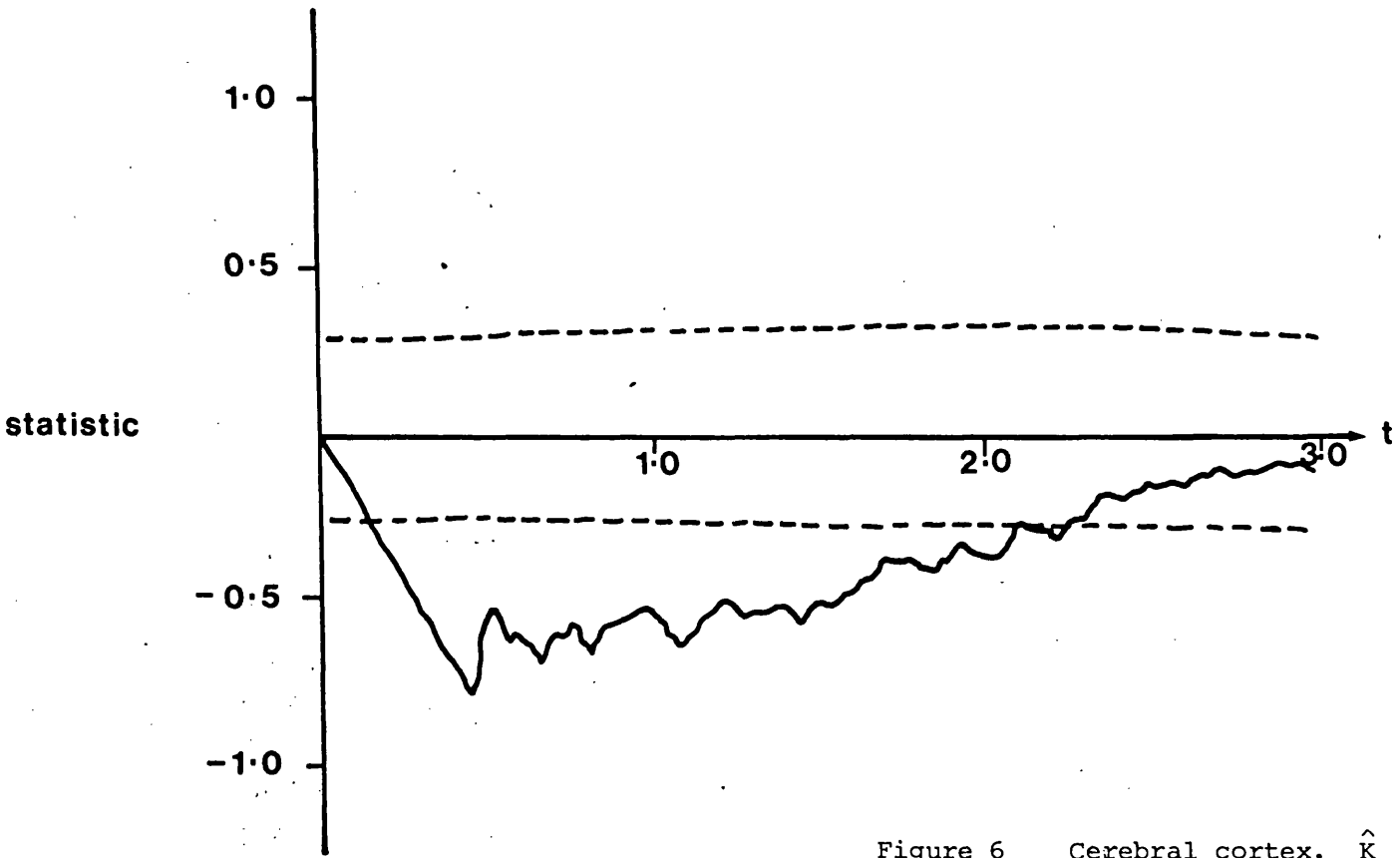


Figure 6 Cerebral cortex. $\hat{K}_{xy}$ for the veins and arteries. The statistic plotted is $\sqrt{\hat{K}_{xy}(t)} - \sqrt{\pi}t$. The dashed lines are approximate 95% confidence bands computed under the double Poisson hypothesis.

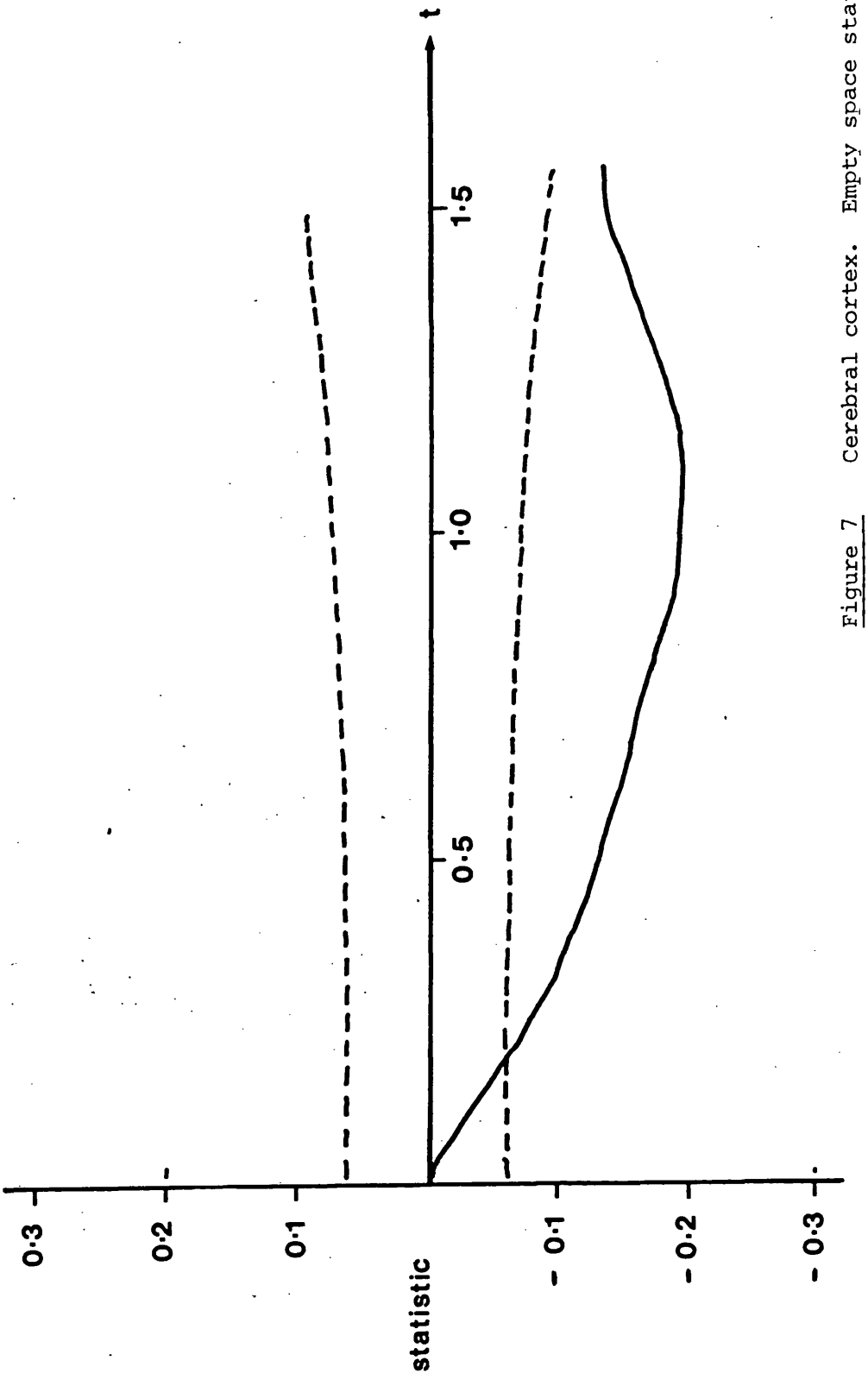


Figure 7 Cerebral cortex. Empty space statistic $T_{xy}(U(t))/t^3$ for the veins and arteries. The dashed lines are asymptotic 95% confidence bands computed under the double Poisson hypothesis.

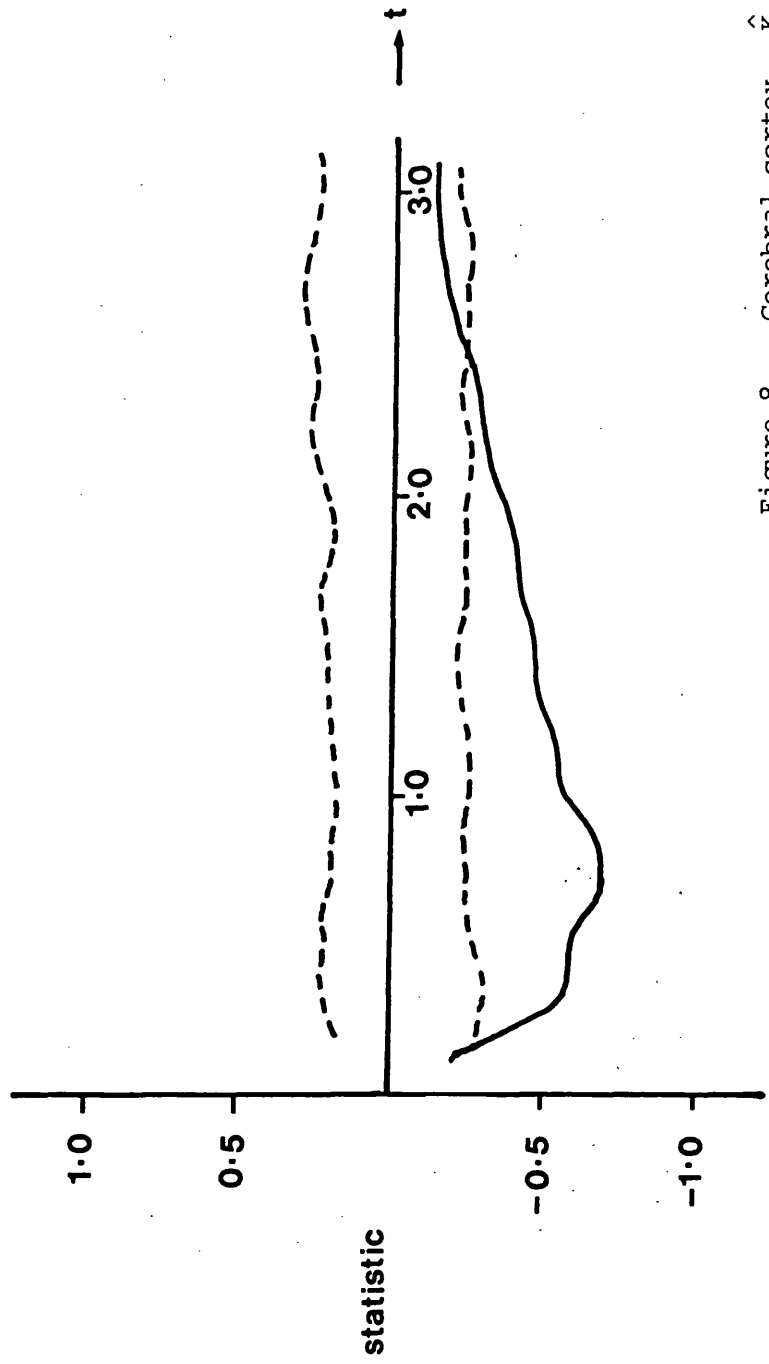


Figure 8 Cerebral cortex. $\hat{K}_{xy}$ on the torus for the veins and arteries. The statistic plotted is $\sqrt{\hat{K}_{xy}(t)} - \sqrt{\pi}t$. The dashed lines are 95% confidence bands obtained from 99 random translations of the torus.

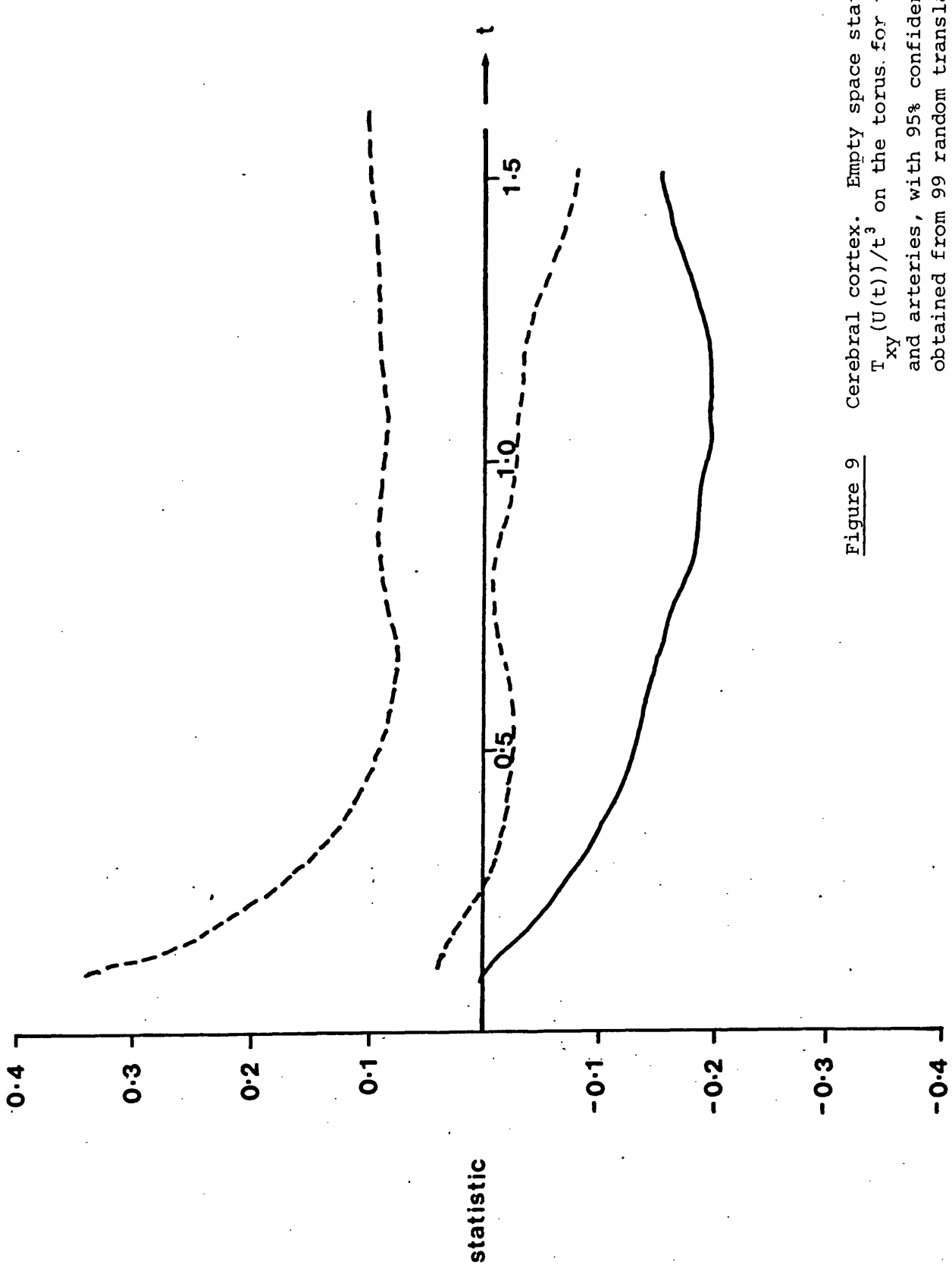


Figure 9 Cerebral cortex. Empty space statistic $T_{xy}(U(t))/t^3$ on the torus for the veins and arteries, with 95% confidence bands obtained from 99 random translations of the torus.

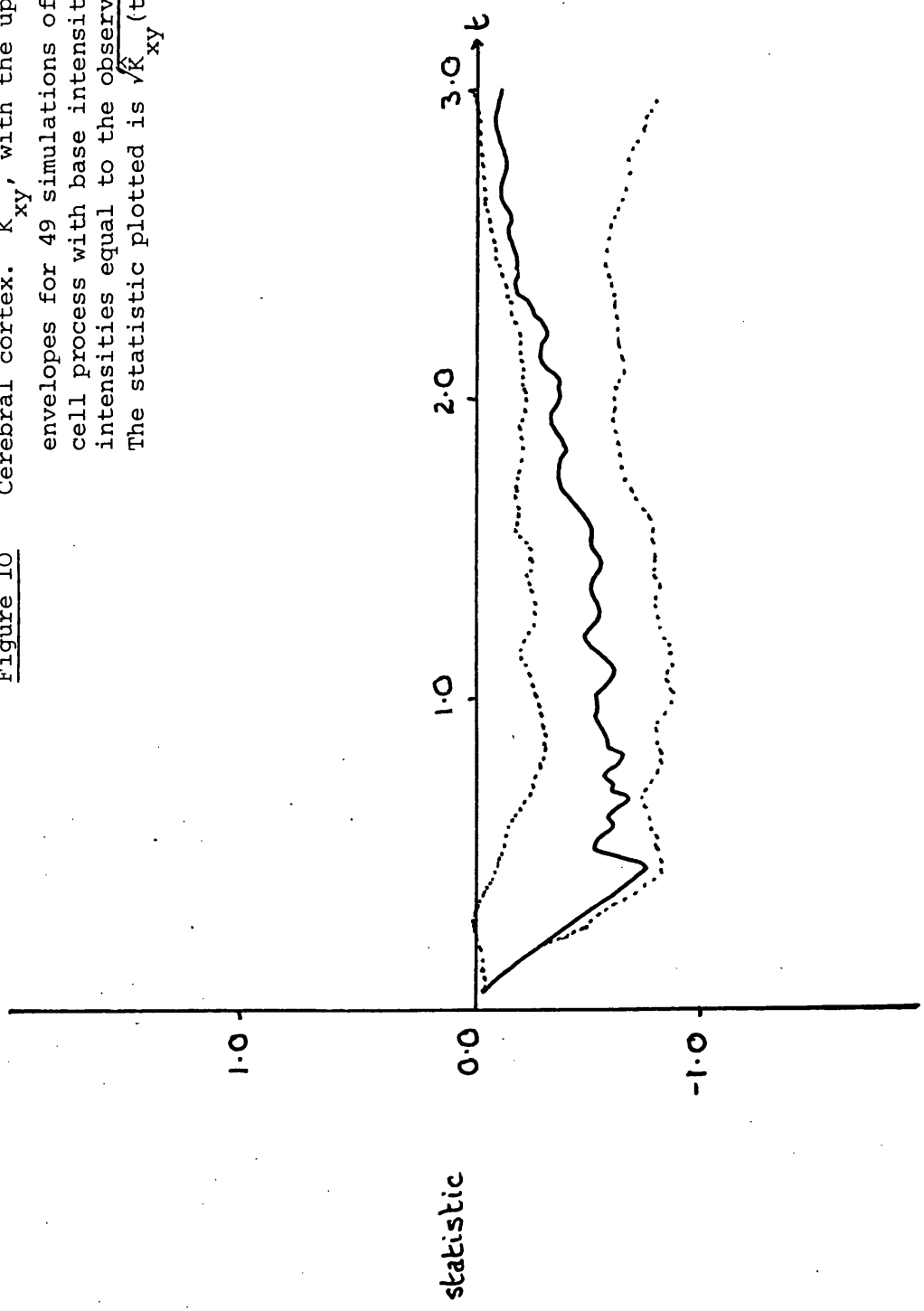


Figure 10 Cerebral cortex. $\hat{K}_{xy}$, with the upper and lower envelopes for 49 simulations of a random Dirichlet cell process with base intensity 0.6, and marginal intensities equal to the observed marginal intensities. The statistic plotted is $\sqrt{\hat{K}_{xy}(t)} - \sqrt{\pi}t$.

The results of the analysis of the two-type pattern using $\hat{K}$ and T are shown in Figures 6 and 7. Both methods indicate considerable inhibition between the two types, enough to reject the hypothesis of independence of the two types of process. Use of conditional Monte Carlo tests as described in section 4 gave similar results. 99 random translations of the torus were performed, and the 95% confidence bands obtained from these are shown in Figures 8 and 9. Both $\hat{K}$ and T were, of course, calculated for a torus in this case.

Thus all the tests used (illustrated in Figures 6, 7, 8 and 9) gave comparable results, showing inhibition between the two types at all but short ranges. Thus the veins 'fill in' the gaps left by the arteries, and vice versa. The fact that the marginal patterns are nearly Poisson suggests the possibility of attempting to fit a bivariate Poisson model of the kind described by Brown, Milne and Silverman (1981). This was done, conditioning on the marginal intensity of the two processes and fitting a random Dirichlet cell process; the results are shown in Figure 10. As can be seen from this figure, a reasonable fit was obtained. The computer program for generating the Brown-Milne-Silverman processes was supplied by Dr. T.C. Brown; it uses the TILE 4 package.

Example 2: Liver

The dataset for this example is shown in Figure 11. It consists of a cross-section through part of a human liver, again with the positions of the veins and arteries marked.

The patterns were analysed in the same way as the data from the cerebral cortex; the results are shown in Figures 12 to 19. This time it is the veins that exhibit some departure from Poisson; this is shown using $\hat{K}$ but not $\tilde{G}$ (see Figures 12 and 14). The arteries can

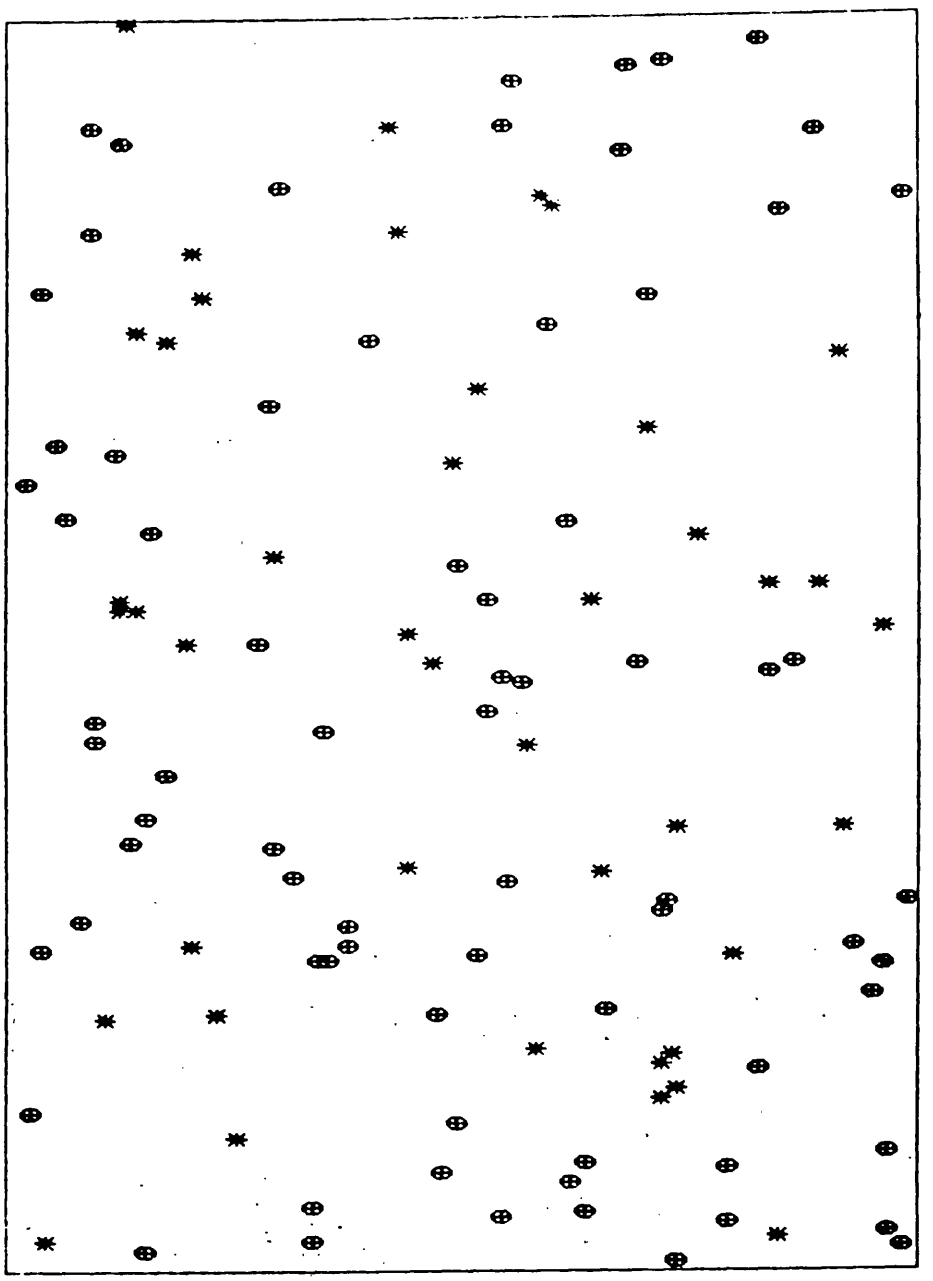


Figure 11 A 25.5 × 18.2 rectangle containing the positions in a cross-section through a human liver of 43 arteries (*) and 74 veins (⊕).

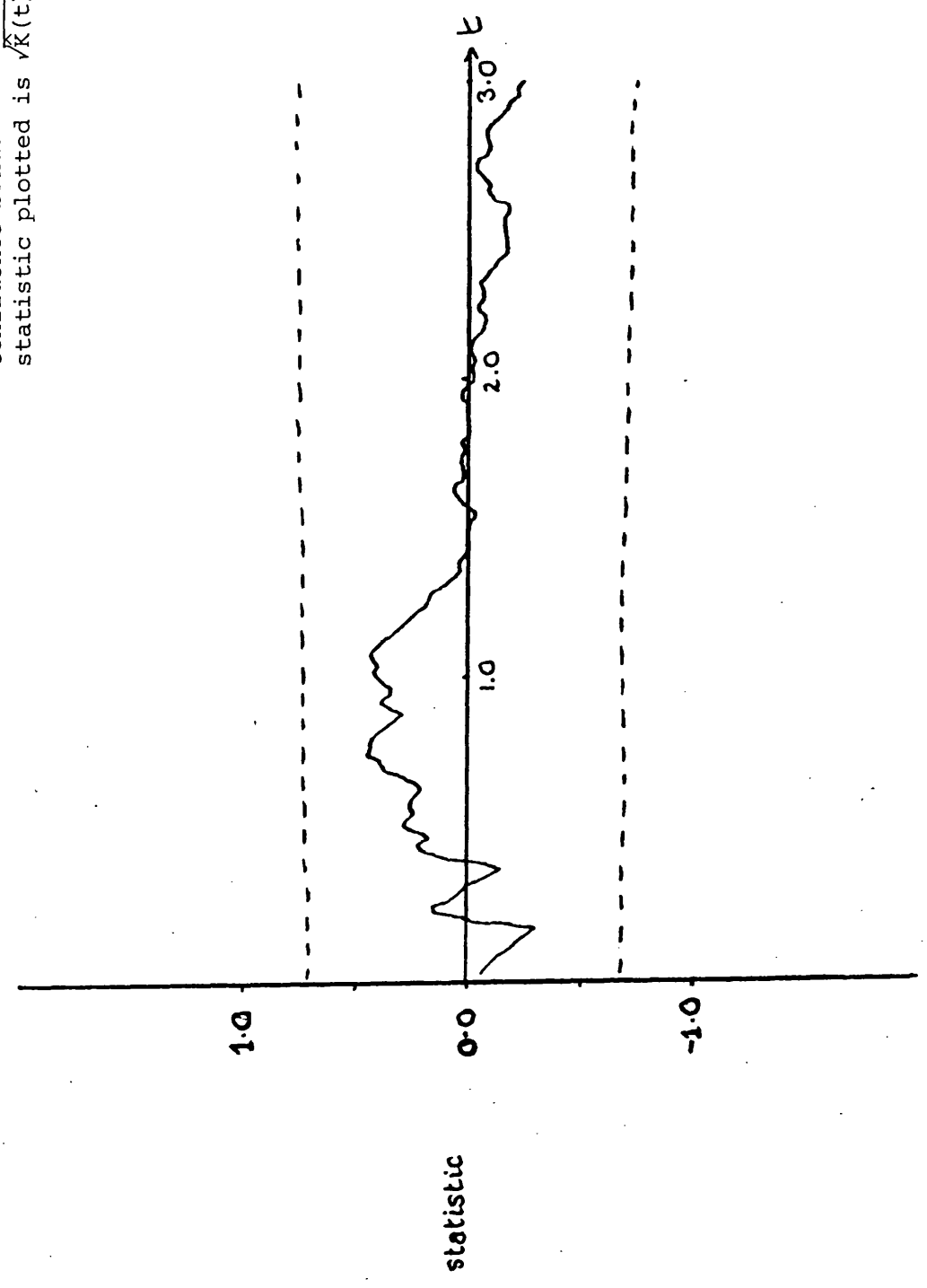


Figure 12 Liver. $\hat{K}$ for the arteries alone, with estimated 95% confidence bands for a Poisson process. The statistic plotted is $\sqrt{\hat{K}(t)} - \sqrt{\pi}t$.

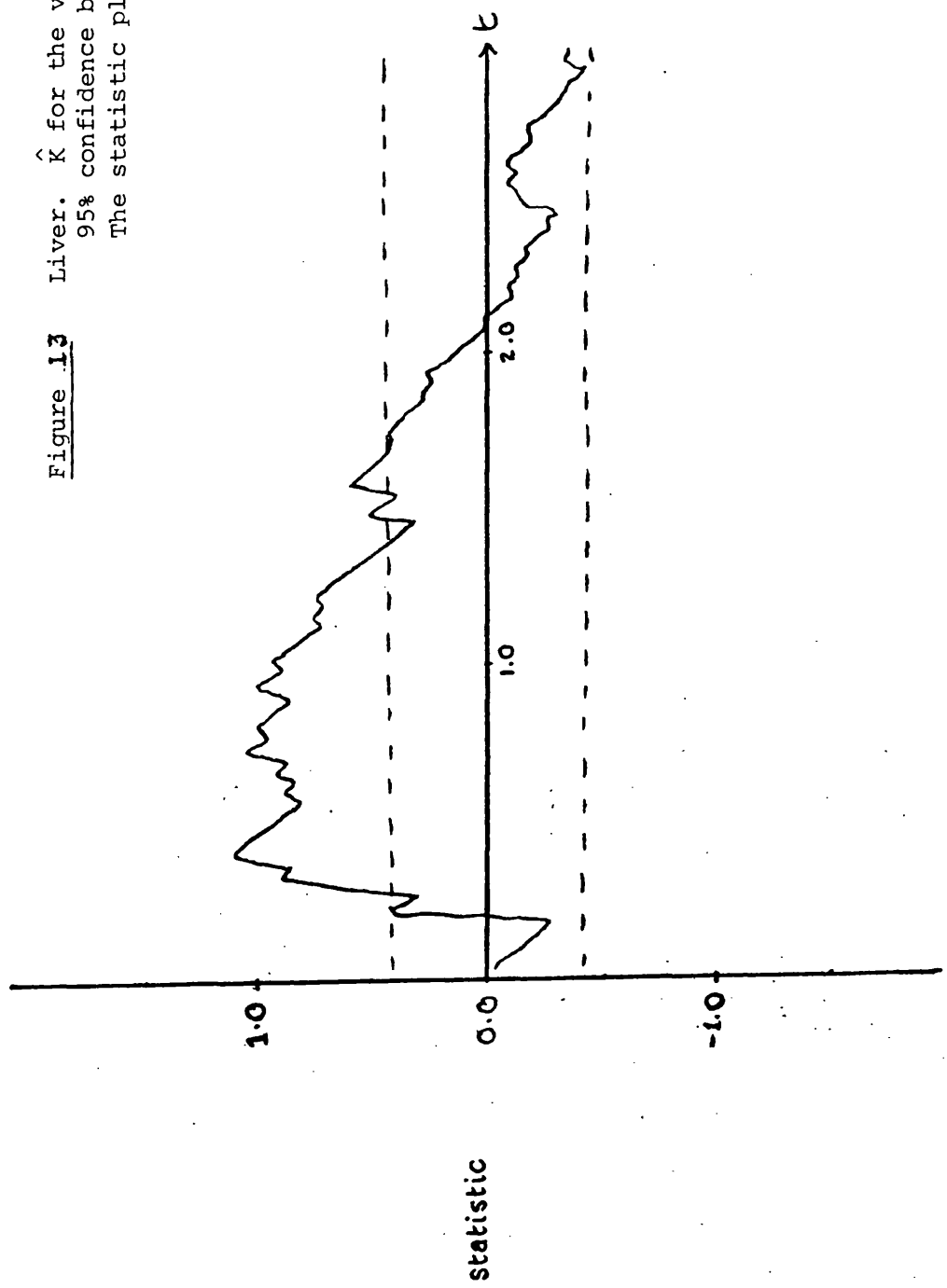


Figure 13 Liver. $\hat{K}$ for the veins alone, with estimated 95% confidence bands for a Poisson process. The statistic plotted is $\sqrt{\hat{K}(t)} - \sqrt{\pi}t$.

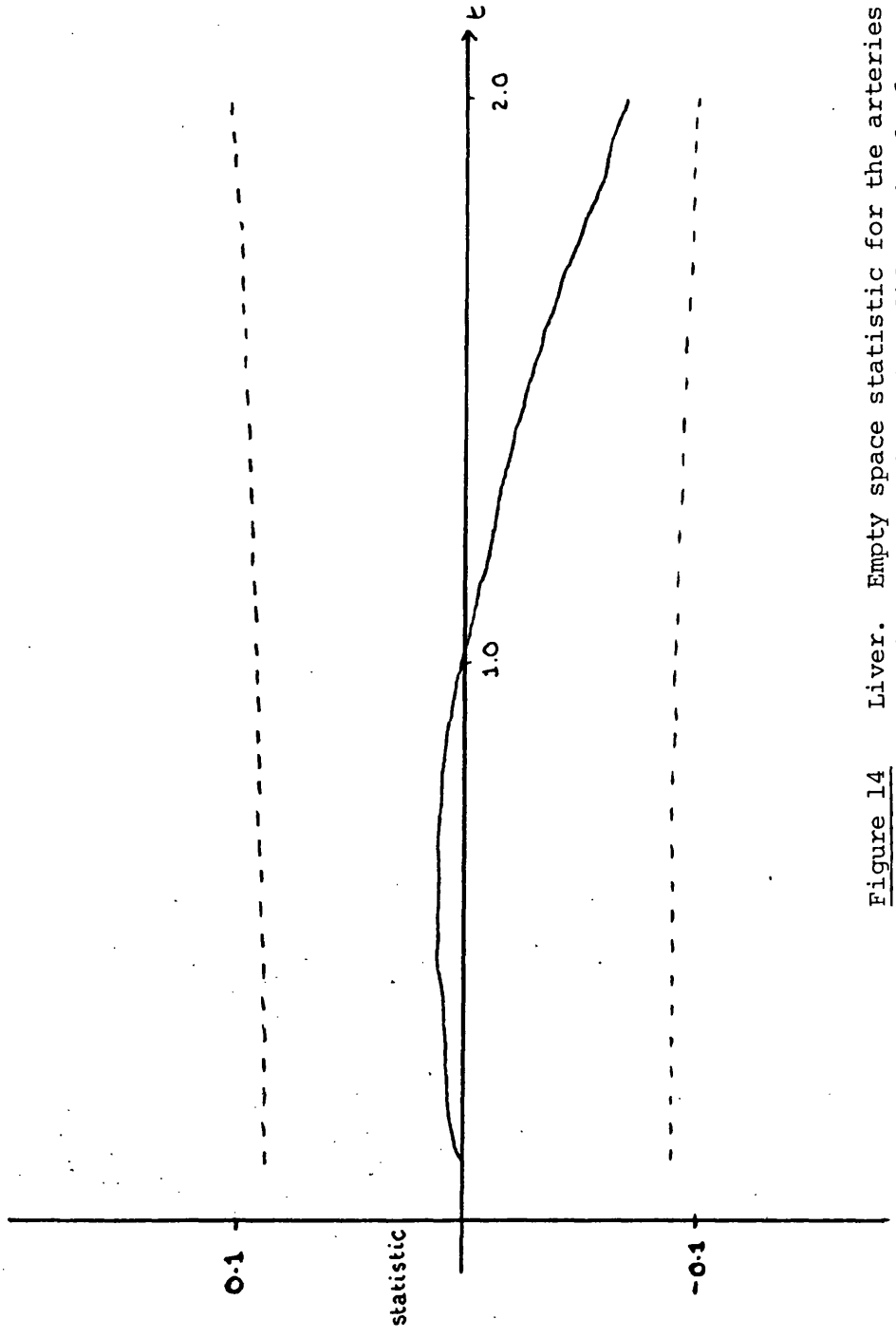


Figure 14 Liver. Empty space statistic for the arteries alone, with asymptotic 95% confidence bands for a Poisson process. The statistic plotted is $\{\log G(U(t)) + .093\pi\}/t^2$.

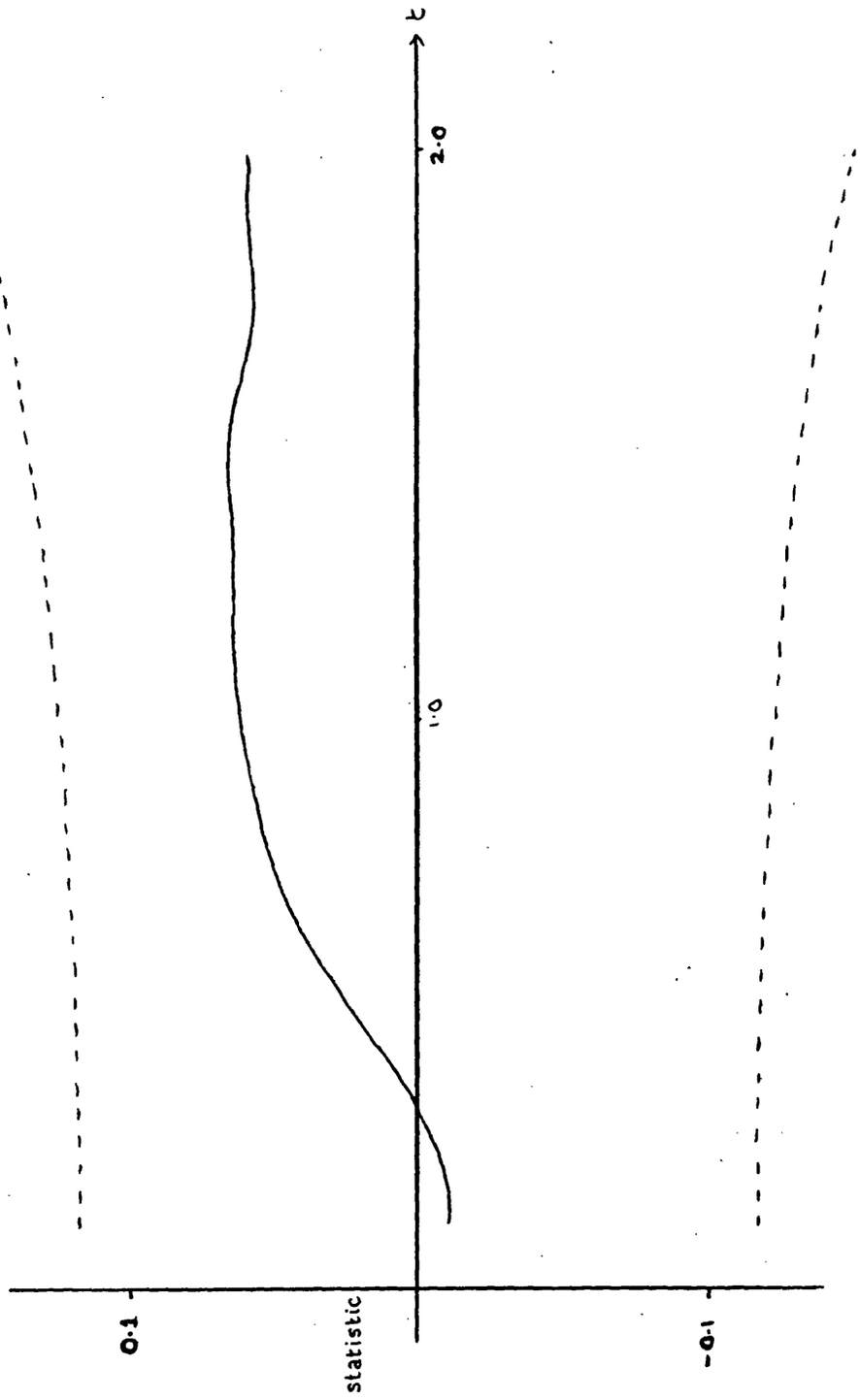


Figure 15 Liver. Empty space statistic for the veins alone, with asymptotic 95% confidence bands for a Poisson process. The statistic plotted is $\{\log \hat{G}(U(t)) + .159\pi\}/t^2$.

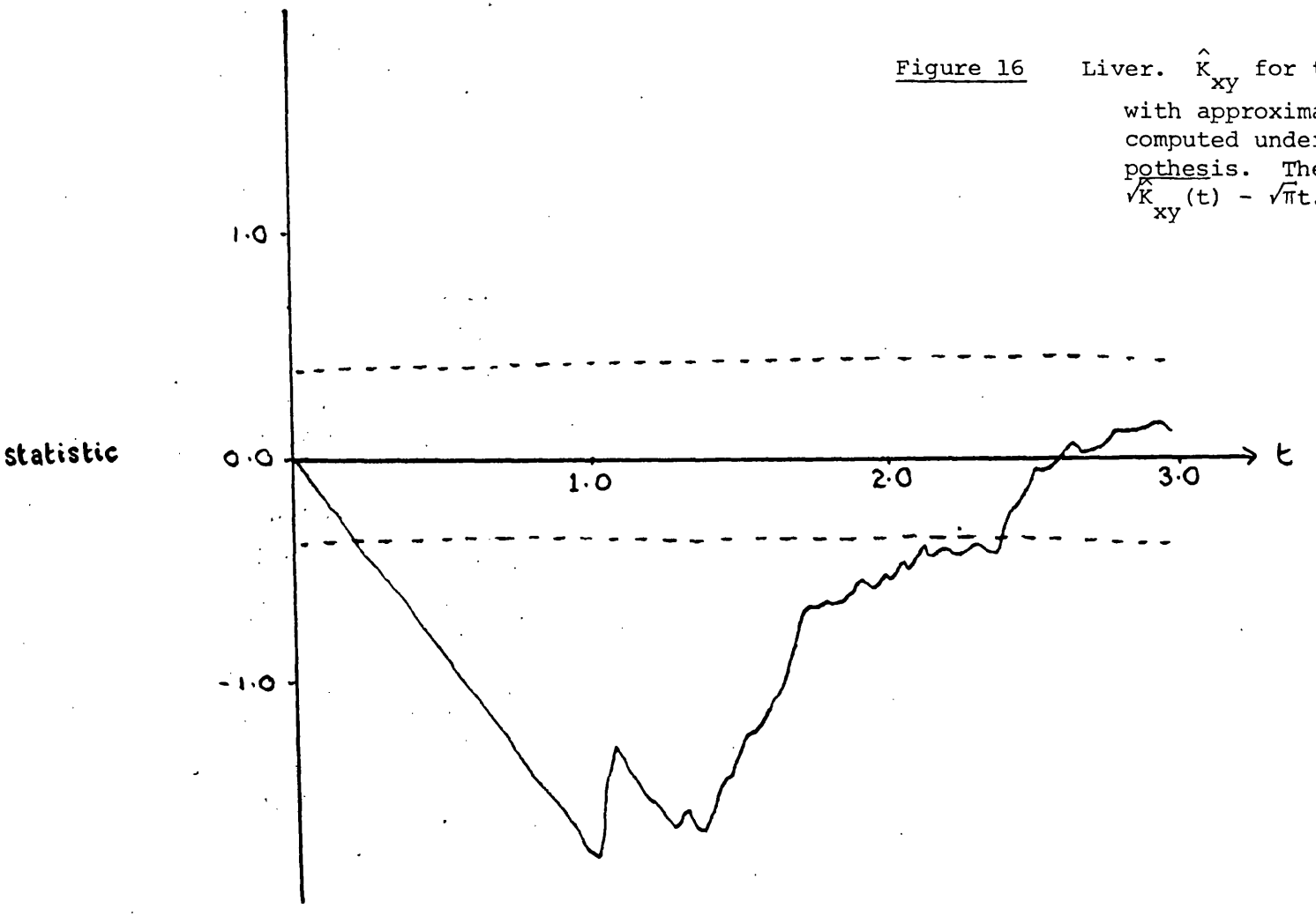


Figure 16 Liver. $\hat{K}_{xy}$ for the veins and arteries, with approximate 95% confidence bands computed under the double Poisson hypothesis. The statistic plotted is $\sqrt{\hat{K}_{xy}(t)} - \sqrt{\pi}t$.

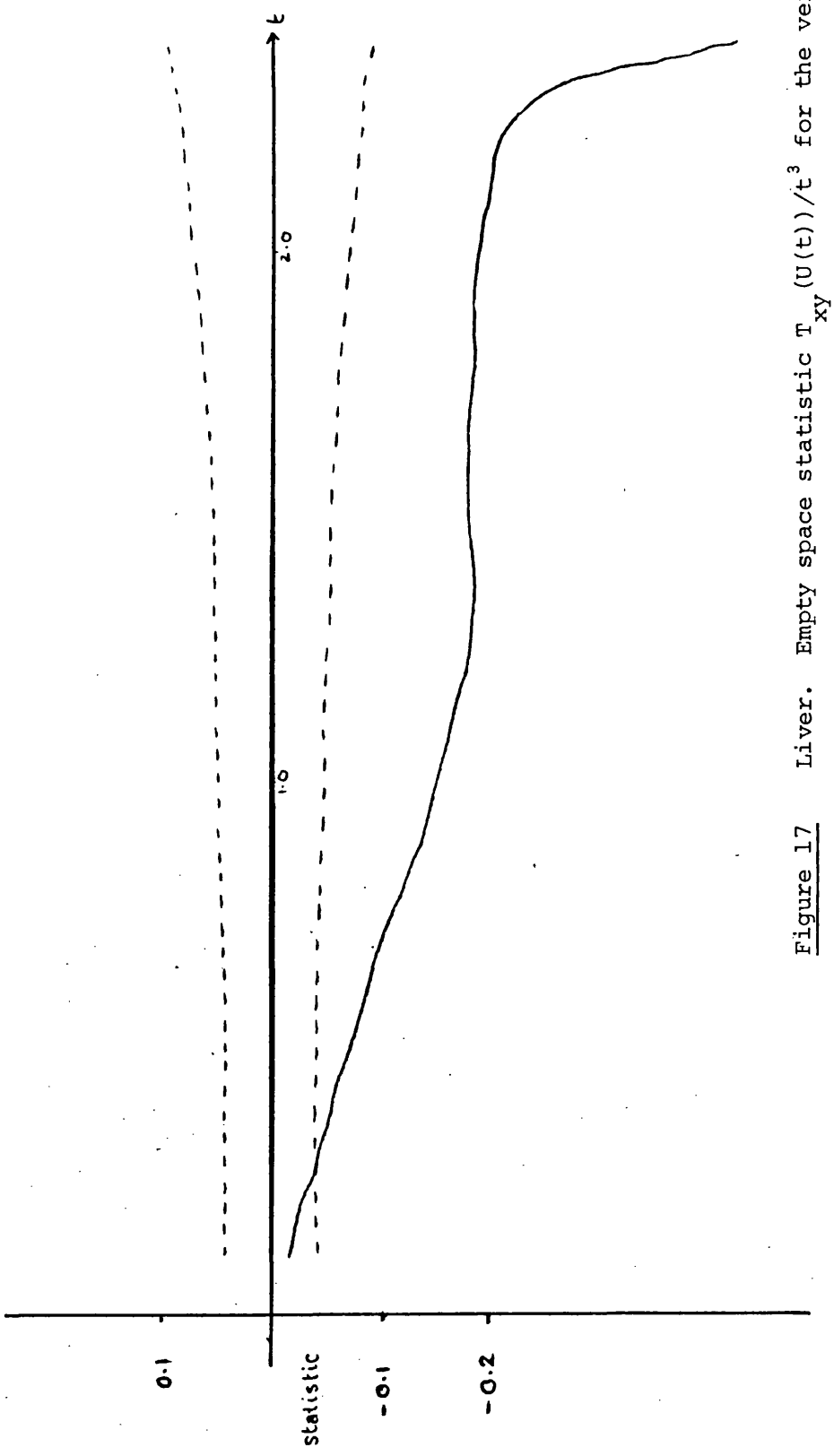


Figure 17 Liver. Empty space statistic $T_{xy}(U(t))/t^3$ for the veins and arteries. The dashed lines are asymptotic 95% confidence bands computed under the double Poisson hypothesis.

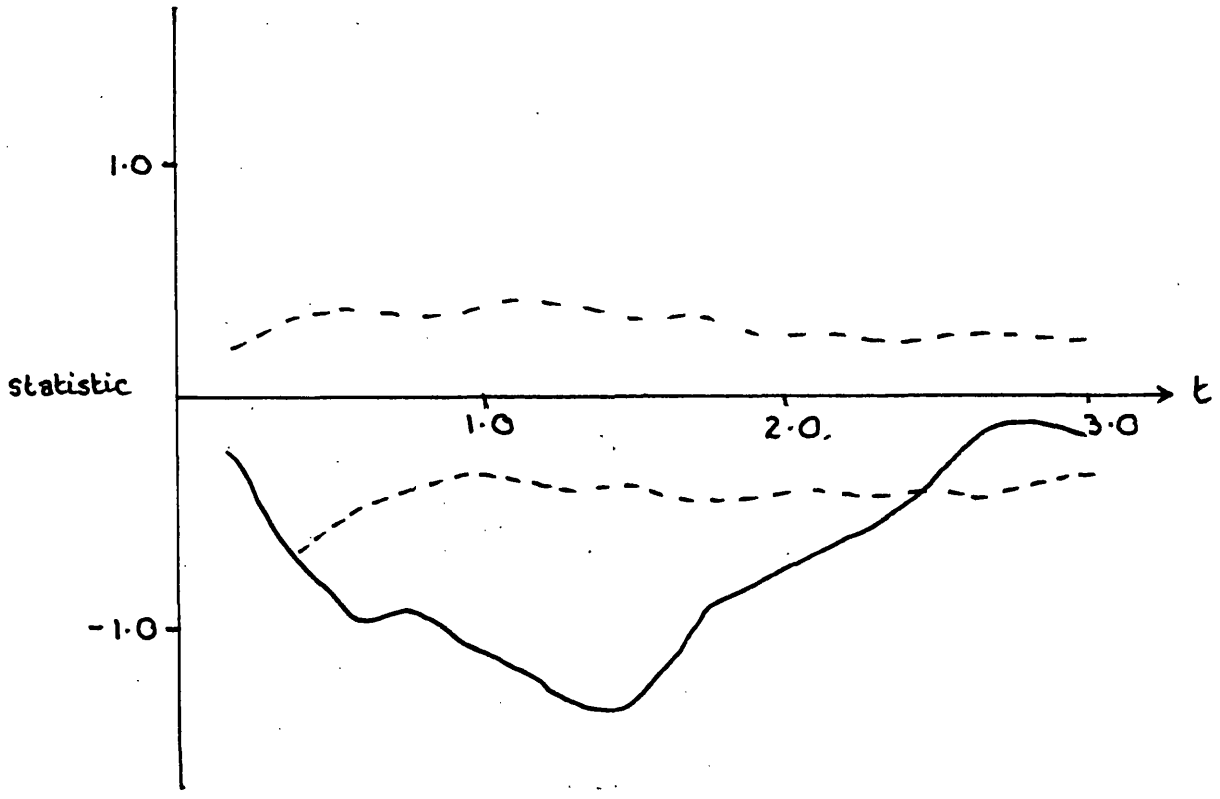


Figure 18 Liver. $\hat{K}_{xy}$ on the torus for the veins and arteries. The statistic plotted is $\sqrt{\hat{K}_{xy}(t)} - \sqrt{\pi}t$. The dashed lines are 95% confidence bands obtained from 99 random translations of the torus.

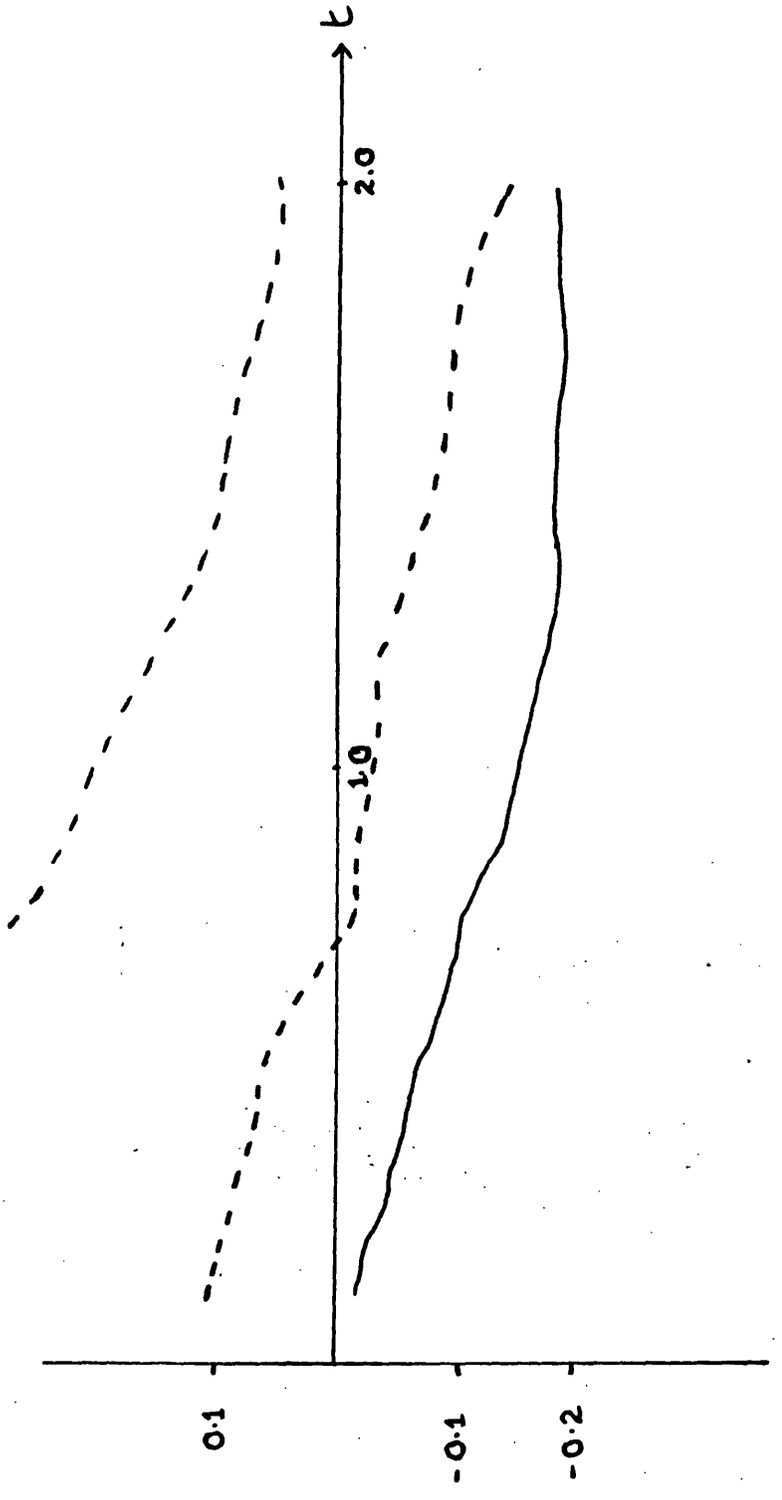


Figure 19 Liver. Empty space statistic $T_{xy}(U(t))/t^3$ on the torus for the veins and arteries, with 95% confidence bands obtained from 99 random translations of the torus.

be accepted as a Poisson process.

Analysing the two-type pattern using asymptotic distributional results and using Monte Carlo methods indicate considerable inhibition between the two types; as in the previous example, this is enough to reject the hypothesis of independence of the two types of process. It is interesting to note that the shapes of the statistic curves in this example are similar to those for the cerebral cortex.

Example 3: Lansing Woods

The pattern for this example (see Figure 20) consists of the positions of two types of trees - 135 black oaks and 514 maples in a square region of Lansing Woods, Michigan. The two single-type patterns were analysed separately in Chapter 1, section 6 and in Chapter 2, section 5; it was seen there that each was well-fitted by a Poisson-Poisson cluster model. Thus, if the two processes are independent, the distributional result of section 3, Theorem 2, will hold. $\hat{K}_{xy}$ and T_{xy} for the data are shown in Figures 21 and 22. Asymptotic 95% confidence bands for the independence hypothesis, calculated by numerical integration on the basis of the fitted parameter values for Poisson-Poisson cluster processes (see section 5 of Chapter 2) have been drawn on the graph of T_{xy}. The confidence bands for $\hat{K}_{xy}$ are calculated on the double Poisson hypothesis, and so cannot be regarded as reliable; it is noted in Silverman and Lotwick (1981) that the bands will be too narrow if the marginal processes display clustering. Both the graphs show marked inhibition between the processes, as do the results using Monte Carlo methods, shown in Figures 23 and 24. This is consistent with the results of Diggle and Milne (1981), who detected negative second order dependence between

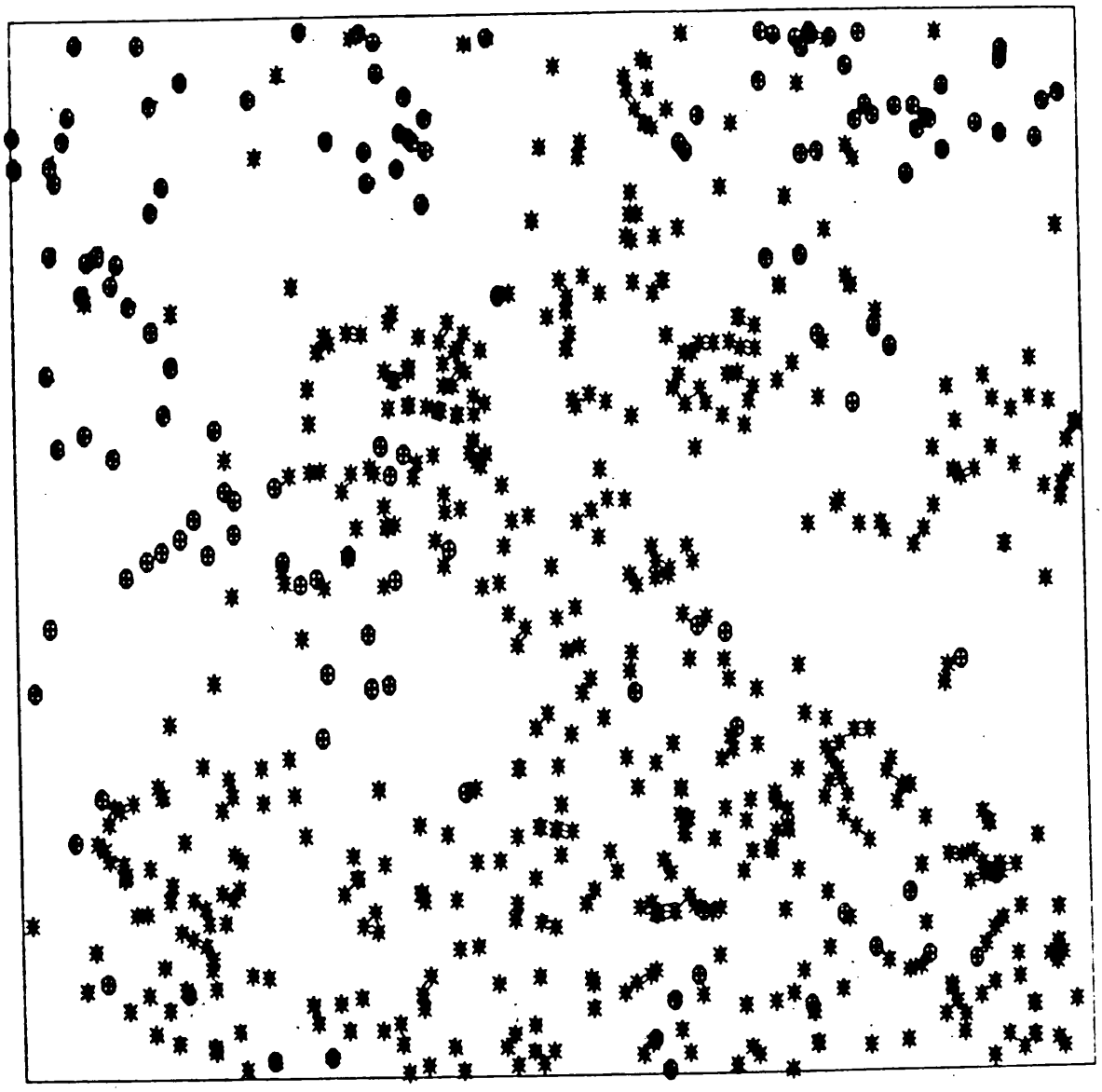


Figure 20 135 black oaks (⊕) and 514 maples (*) in a 924 ft × 924 ft region of Lansing Woods

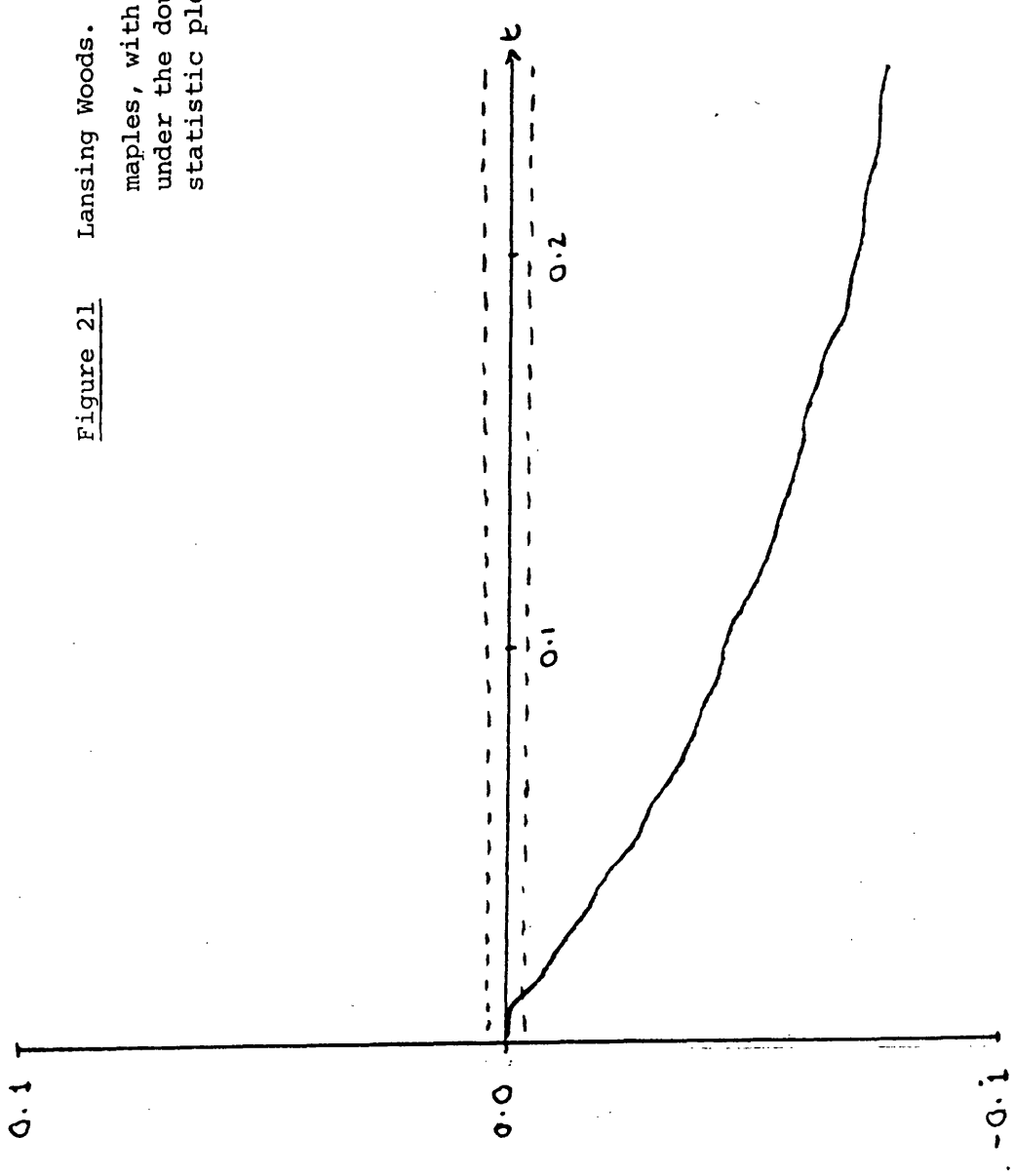


Figure 21 Lansing Woods. $\hat{K}_{xy}$ for the black oaks and maples, with 95% confidence bands computed under the double Poisson hypothesis. The statistic plotted is $\sqrt{\hat{K}_{xy}(t)} - \sqrt{\pi}t$.

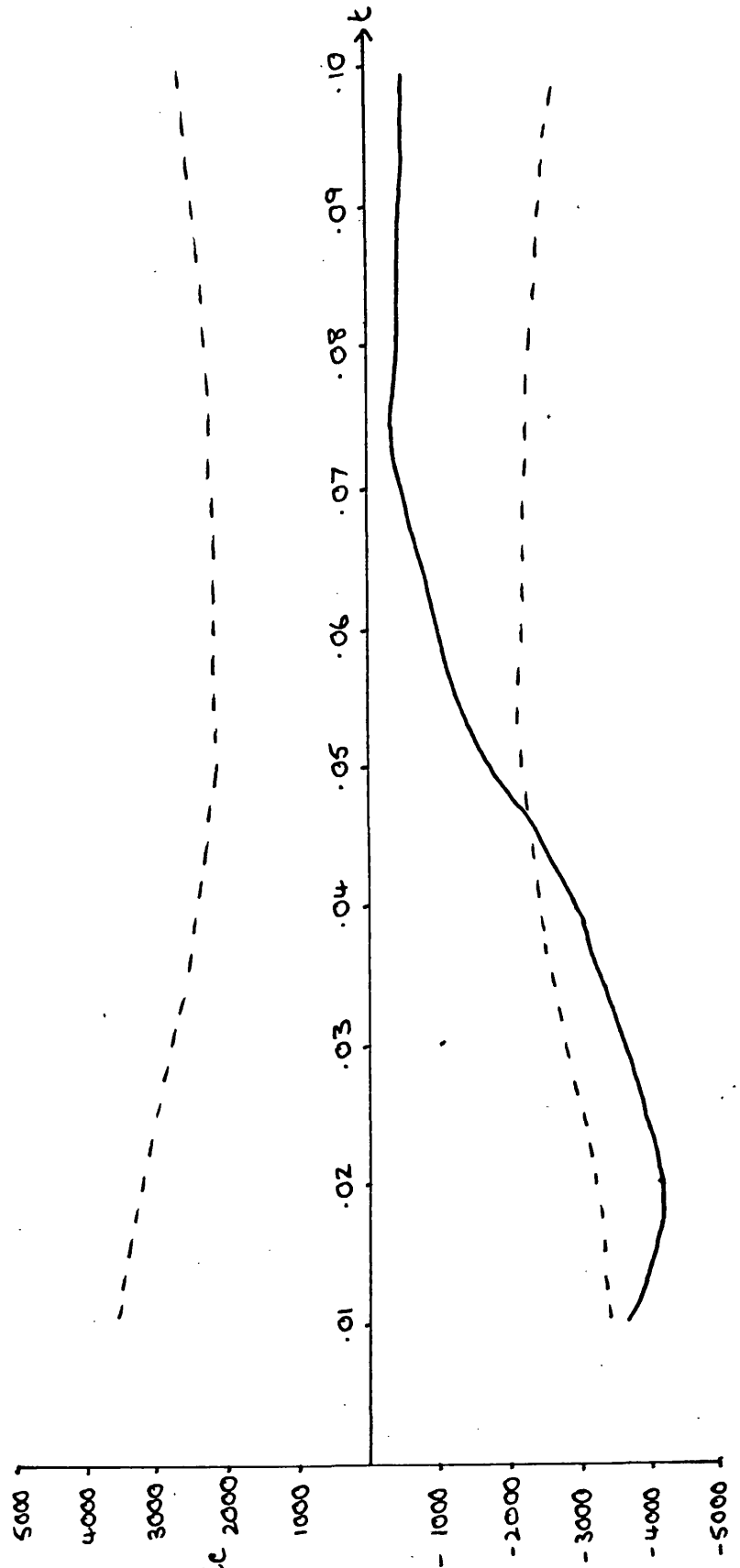


Figure 22 Lansing Woods. The empty space statistic $T_{xy}(U(t))/t^3$ plotted for the black oaks and maples. The dashed lines are 95% confidence bands under the hypothesis of independence, using the Poisson-Poisson cluster processes fitted in Chapter 2 as the marginal distributions.

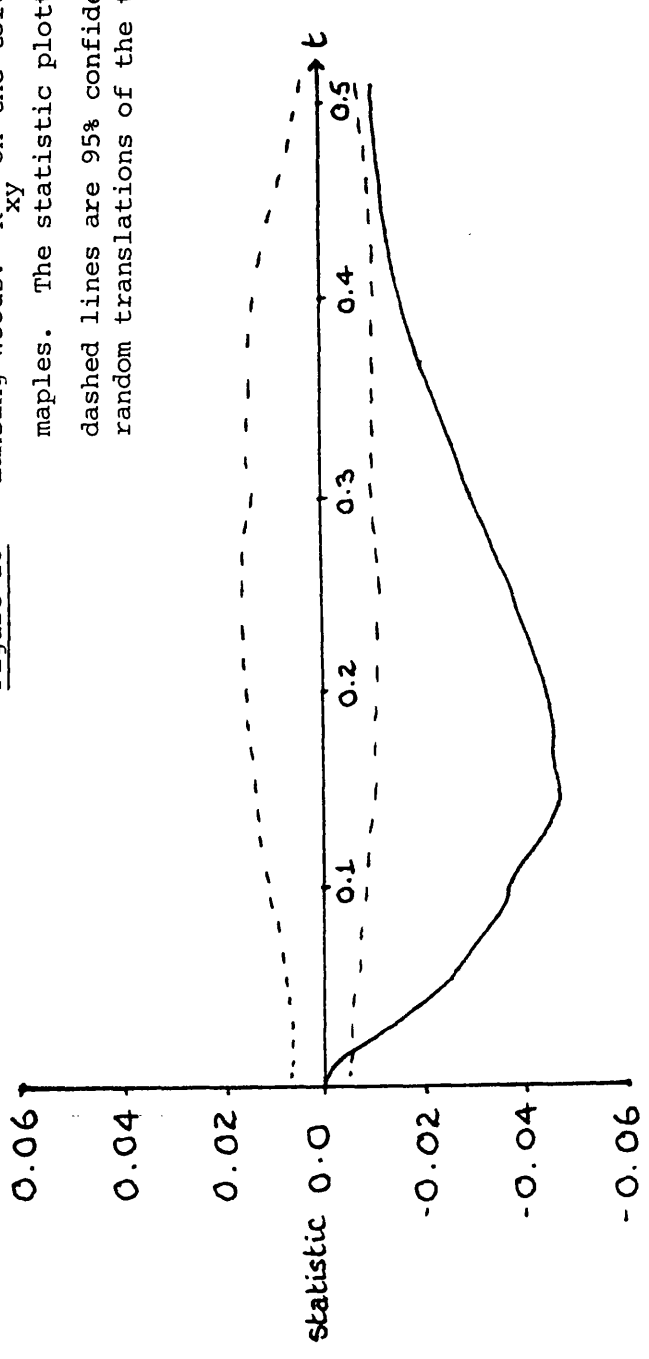


Figure 23 Lansing Woods. $\hat{K}_{xy}$ on the torus for the black oaks and maples. The statistic plotted is $\sqrt{\hat{K}_{xy}(t)} - \sqrt{\pi}t$. The dashed lines are 95% confidence bands obtained from 99 random translations of the torus.

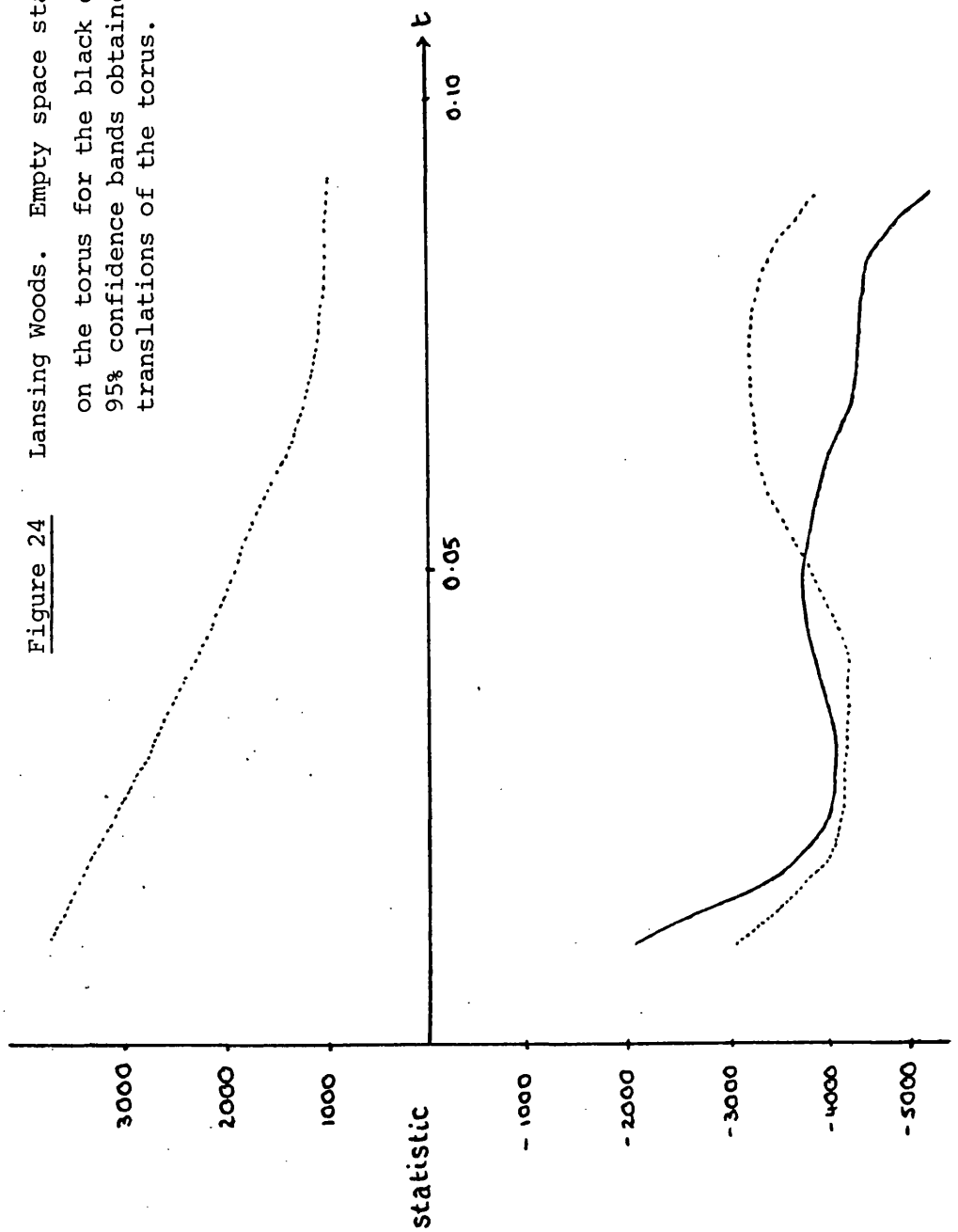


Figure 24 Lansing Woods. Empty space statistic $T_{xy}(U(t))/t^3$ on the torus for the black oaks and maples, with 95% confidence bands obtained from 99 random translations of the torus.

Lansing Woods maples and oaks (of which the black oaks form a part).

Comment

As the three examples of this section show, empty space methods provided a very useful tool for analysing multitype point patterns. They can be viewed as being complementary to Silverman's distance methods; the two techniques should be used side by side in the analysis of a multitype pattern, giving greater insight into the behaviour of the processes.

The hypothesis of independence of the two processes can be viewed as a 'dividing hypothesis' in the sense of Cox (1977), rather than one to which credence is attached; it thus serves as a reference line for dividing two-type processes into ones in which the types attract each other, and ones which display inhibition between the types.

6. Extensions

All the discussion in the previous sections has been concerned with two-type point processes. Whereas for second-order methods, a pattern with more than two types is analysed by considering the types in pairs (see section 2.1 of Silverman and Lotwick (1981)), empty space methods make it possible to consider such patterns as a whole, and look for general, rather than two-way, interactions between the types. For example, if we have three types of point, then, if we have total independence between the three types (X-, Y- and Z-, say), we have, analogously to (3.2), for any Borel subset A of the plane,

$$\log G_{X\cup Y\cup Z}(A) = \log G_X(A) + \log G_Y(A) + \log G_Z(A) \qquad (3.8)$$

More generally, though, we can write

$$\begin{aligned}\log G_{X\cup Y\cup Z}(A) = {} & \log G_X(A) + \log G_Y(A) + \log G_Z(A) \\ & + \alpha_{XY}(A) + \beta_{YZ}(A) + \gamma_{XZ}(A) \\ & + \delta_{XYZ}(A) \qquad (3.9)\end{aligned}$$

where

$$\alpha_{XY}(A) = \log G_{X\cup Y}(A) - \log G_X(A) - \log G_Y(A)$$

$$\beta_{YZ}(A) = \log G_{Y\cup Z}(A) - \log G_Y(A) - \log G_Z(A)$$

$$\gamma_{XZ}(A) = \log G_{X\cup Z}(A) - \log G_X(A) - \log G_Z(A)$$

and

$$\begin{aligned}\delta_{XYZ}(A) = {} & \log G_{X\cup Y\cup Z}(A) - \log G_{X\cup Y}(A) - \log G_{Y\cup Z}(A) \\ & - \log G_{X\cup Z}(A) + \log G_X(A) + \log G_Y(A) + \log G_Z(A) \,.\end{aligned}$$

Here, α, β and γ are respectively measures of the X-Y, Y-Z and X-Z interactions, and δ is a measure of the interaction between the three

types not accounted for by the two-way interactions. Each of these quantities can be estimated from a point pattern in the obvious way, thus providing insight into the behaviour of the process. These remarks extend naturally to patterns with more than three types of point.

CHAPTER FOUR

HARD CORE MODELS

1. Introduction

A hard core point pattern is one in which the points of the pattern may be regarded as being the centres of non-overlapping discs of fixed diameter R. The condition that the discs be non-overlapping is equivalent to requiring that none of the inter-point distances is less than R; a hard core process is thus an extreme case of a 'regular' or inhibited process. (The use of R as a symbol for the diameter of the discs tallies with the notation used for Kelly-Ripley models (see Chapter 5)).

Although the discussion in this chapter will be in terms of two-dimensional hard core patterns, much of the work is easily generalised to three dimensions, where the points are the centres of non-overlapping spheres. Models for patterns of non-overlapping discs or spheres are used in various branches of science, including Mechanics of Random Media (Talbot and Willis (1980), Willis (1980)), Chemical Physics (Alder and Wainwright (1962)) and Haematology (Mohn and Stavem (1974)).

The packing density of a hard core process or pattern is the proportion of the plane covered by the discs. A hard core process of intensity λ thus has packing density $\lambda\pi R^2/4$, since the discs have area $\pi R^2/4$. The theoretical bound on the packing density of a hard core process on an infinite region is $\pi/\sqrt{12}$ (0.906), which is attained by a triangular lattice packing (see Rogers, 1964).

In this chapter, various models for hard core patterns are described, and some of their properties are studied. In the following chapters, some specific problems concerned with hard core processes are investigated.

2. Models for hard core patterns

Various models for hard core two-dimensional patterns have been proposed. All the models below can be adapted to give three-dimensional models.

i) Matérn Model 1 (Matérn, 1960): This is obtained by sampling a Poisson process of intensity λ and deleting any point which is within R of any other, whether or not this other point has already been deleted. The resulting process is hard core, with intensity $\lambda \exp(-\pi\lambda R^2)$; the model has maximum packing density, as λ is varied, of $(4e)^{-1}$.

ii) Matérn Model 2 (Matérn, 1960): Here, the points of a Poisson process of intensity λ are independently marked with a uniformly distributed birth time on (0,1); a point is deleted if there is within distance R of it a point of the Poisson process having an earlier birth time. The upper bound on the packing density of such a process is 0.25.

Some properties of the Matérn models are discussed in section 4 of this chapter.

iii) Mürmann's Exclusion Model. Mürmann (1978) defined Poisson point processes with exclusion by their local conditional distributions. Use of an exclusion set consisting of a disc of diameter R centred at the point in question produces, as Mürmann notes in his Introduction, a model for configurations of non-overlapping discs with this diameter. Since the models are defined in terms of conditional distributions, the existence and uniqueness of these processes needs to be proved. Mürmann (1978) proved existence, but for an unbounded sampling region, the uniqueness problem remains open. Mürmann (1978) conjectures that

uniqueness is violated if the packing density is high enough.

Each of the three hard core models above is reformulated in terms of random measures in section 3 of this chapter; such a reformulation makes the Mürmann process easier to handle.

Models i) - iii) above can be used to give hard core patterns on the whole of $\mathbb{R}^2$. The following hard core models can be defined on a compact sampling region E:

iv) The Kelly-Ripley hard core model. This is a special case (c = 0) of a wider class of models for inhibited point processes defined by Kelly and Ripley (1976). It may be described as follows: a Poisson process is sampled on E, and only those samples containing no pair of points less than R apart are retained. This model is, of course, the same as model iii) above defined on E rather than $\mathbb{R}^2$. It may be used to model configurations with any packing density up to the theoretical limit. Some problems relating to processes used to simulate Kelly-Ripley models are treated in Chapters Five and Six.

v) The SSI Model: The Simple Sequential Inhibition (SSI) process was introduced by Diggle, Besag and Gleaves (1976). Discs are placed sequentially with their centres in E, successive disc centres being chosen uniformly from those points which ensure that no two discs overlap; the procedure terminates when a prescribed number N of discs have been inserted. Simulation of this process is discussed in Chapter Seven, where an empirical study is made of the packing densities attainable using the process.

Various other models have been proposed for hard core patterns, including ones where the discs are of variable size (see Ripley, 1977, section 3.2); we shall restrict ourselves to models i) - v) above.

Of these models listed above, i), ii) iv) and v) are the most commonly used. Note that in each of the models except model iv), the packing density has an upper bound which is lower than the maximum packing density for hard core patterns.

3. Random Measure Formulation of Some Hard Core Models

A point process in the plane can be viewed as a random element from the space of locally finite non-negative integer-valued measures on the Borel sets of $\mathbb{R}^2$; here, the measure of a set B is the number of points of the process lying in B. Random measures are discussed in detail by Kallenberg (1975). In this section, we reformulate three of the hard core models of section 2 in terms of random measures. Such a formulation makes the Mürmann model somewhat more tractable, and enables us to put the unresolved uniqueness problem for the model in a different form.

We use the notation of Kallenberg. For $x \in \mathbb{R}^2$, write

$$I(x) = \{y \in \mathbb{R}^2 : |x - y| \leqslant R\} . \qquad (4.1)$$

Denote by η the class of all locally finite measures on $(\mathbb{R}^2, \mathcal{B}^2)$ taking values in $\mathbb{Z}_+$; here $\mathcal{B}^2$ is the collection of Borel subsets of $\mathbb{R}^2$. We restrict to point processes which have no coincident points ('simple' point processes in Kallenberg's terminology), writing

$$\eta^* = \{\mu \in \eta : \mu \text{ simple}\}.$$

Let

$$\eta_o = \{\mu \in \eta^* : \mu(\{x\} \cup \{y\}) \leqslant 1 \text{ if } x \in I(y)\} \qquad (4.2)$$

and let N_o be the σ-algebra in η_o generated by the mappings $\mu \to \mu B$, $B \in \mathcal{B}^2$.

Definition A hard core point process is a probability measure on (η_o, N_o).

i) Random Measure Formulation of Matérn Model 1

If $\mu \in \mathcal{N}^*$, then we can (Kallenberg, section 2) decompose μ uniquely (apart from the order of the terms) as

$$\mu = \sum_{j=1}^{\infty} \delta_{t_j}$$

where $\{t_1, \ldots, t_j, \ldots\}$ is a measurable function of μ, and $\delta_s B = 1_B(s)$, $B \in \mathcal{B}^2$.

Define

$$h(t_j,t_k) = 1\{t_j \notin I(t_k)\} .$$

For A compact, $A \subseteq \mathbb{R}^2$, write

$$\mu_A^\dagger = \sum_{t_j \in A} \{ \prod_{t_k \in A} h(t_j,t_k)\}\delta_{t_j} + \sum_{t_j \notin A} \delta_{t_j} ;$$

i.e. we delete pairs of atoms of μ within R of each other which lie in A.

Define $g_A^\dagger : \mathcal{N}^* \to \mathcal{N}^*$ by

$$g_A^\dagger(\mu) = 1_A \mu_A^\dagger .$$

The pointwise product $1_A \mu_A^\dagger$ is an element of $\mathcal{N}_o$, so $g_A^\dagger$ is in fact a function from $\mathcal{N}^*$ to $\mathcal{N}_o$. Since I(x) is measurable, $g_A^\dagger$ is a measurable function. A probability measure on $\mathcal{N}^*$ induces a probability measure on $\mathcal{N}_o$, via $g_A^\dagger$.

Now, take $A_n \uparrow \mathbb{R}^2$, with A_n compact, and such that $\text{int}(A_n) \uparrow \mathbb{R}^2$. Let η be a Poisson process on $\mathbb{R}^2$, and let ξ_n be the point processes induced by $g_{A_n}^\dagger$.

Let $f \in \mathcal{F}_c$, the class of continuous $\mathcal{B}^2$ - measurable functions with compact support. Then $B = \text{supp}(f)$ is compact, and $\exists\ n_o$ s t. $B \subset A_{n_o}$, so that $\xi_n f = \xi_m f$ for $n,m \geq n_o$. Hence $\xi_n f \xrightarrow{d}$ some ξ_f, for all $f \in \mathcal{F}_c$. (We use throughout this section the topology induced on $\mathcal{N}^*$ by the vague topology on $\mathcal{N}$ (see Kallenberg, 1975).) Hence, by Lemma 5.1 of Kallenberg (1975), $\xi_n \xrightarrow{d}$ some ξ.

Since the set of hard core measures is closed, ξ must be a hard core point process. This is easily seen to be the Matérn 1 hard core process.

ii) Random Measure Formulation of the Matérn Model 2

A similar procedure to that above can be used to produce the Matérn 2 hard core process.

For this formulation, we need to consider random measures on $\mathbb{R}^2 \times (0,1)$. We use the product topology on $\mathbb{R}^2 \times (0,1)$, and denote by $\mathcal{C}$ the Borel sets of $\mathbb{R}^2 \times (0,1)$. Denote by $\mathcal{P}^*$ the class of simple locally finite integer-valued measures on $(\mathbb{R}^2 \times (0,1), \mathcal{C})$, and let $\mu \in \mathcal{P}^*$. Then we can write

$$\mu = \sum_{j=1}^{\infty} \delta_{y_j}$$

where, for each j, $y_j = (x_j,t_j)$, with $x_j \in \mathbb{R}^2$ and $t_j \in (0,1)$. Define

$$h(y_j,y_k) = 1 - (1 - 1(t_j < t_k))(1 - 1(x_j \notin I(x_k)))$$

Now, for A compact, $A \subset \mathbb{R}^2$, generate from μ a measure μ_A^* on $(\mathbb{R}^2,\mathcal{B}^2)$ defined by

$$\mu_A^* = \sum_{y_j:x_j \in A} \left\{ \prod_{y_k:x_k \in A} h(y_j,y_k) \right\} \delta_{x_j} + \sum_{y_j:x_j \notin A} \delta_{x_j}$$

Again, define $g_A^* : \mathcal{P}^* \to \mathcal{N}^*$ by

$$g_A^*(\mu) = 1_A \mu_A^* \quad .$$

The pointwise product $1_A \mu_A^*$ is an element of $\mathcal{N}_o$, so g_A^* is a function from $\mathcal{P}^*$ to $\mathcal{N}_o$; it is easily seen to be measurable. A probability measure on $\mathcal{P}^*$ induces a probability measure on $\mathcal{N}_o$, via g_A^*.

A point process on $(\mathbb{R}^2 \times (0,1), C)$ is simply a probability measure on $(\mathcal{P}^*, \mathcal{P}^*)$, where $\mathcal{P}^*$ is the σ-algebra in $\mathcal{P}^*$ generated by the mappings $\mu \to \mu C$, $C \in C$, and a Poisson process on $\mathbb{R}^2 \times (0,1)$ is easily defined.

Take $A_n \uparrow \mathbb{R}^2$, with A_n compact and such that $\text{int}(A_n) \uparrow \mathbb{R}^2$. Let η be a Poisson process on $\mathbb{R}^2 \times (0,1)$, and let ξ_n be the point processes induced by $g_{A_n}^*$. Let $f \in \mathcal{F}_c$; then $B = \text{supp}(f)$ is compact, and $\exists\ n_o$ s.t. $B \subseteq A_{n_o}$, so that $\xi_n f = \xi_m f$ for $n,m \geq n_o$. Hence $\xi_n f \xrightarrow{d}$ some ξ_f and so, by Lemma 5.1 of Kallenberg (1975),

$$\xi_n \xrightarrow{d} \text{some } \xi.$$

ξ is a hard core point process, which is easily seen to be the Matérn 2 hard core process; t_j is the 'birth time' of the point at x_j, so that, in applying g_A^*, we delete all points in A that are within distance R of another point in A having an earlier birth time.

The above formulations of hard core processes could be made with definitions of I(x) other than that of (4.1), giving general 'exclusion' processes. This should be borne in mind in the following paragraphs, where we give an alternative derivative of Mürmann's Poisson point process with exclusion. The only assumptions made about I are that $I(A) \triangleq \bigcup_{x \in A} I(x)$ is a Borel set for every Borel set A of $\mathbb{R}^2$, and that I(U) is bounded if U is bounded.

iii) Random Measure Formulation of Mürmann Model

We first define a Poisson process with exclusion on a compact subset A of $\mathbb{R}^2$.

Define $\eta_o^A \subseteq \eta_o$ by

$$\mu \in \eta_o^A \Leftrightarrow x \in A,\ y \in A,\ y \in I(x) \Rightarrow \mu(\{x\} \cup \{y\}) \leqslant 1. \qquad (4.3)$$

Let η be a Poisson process on $\mathbb{R}^2$. Define a new point process $\xi^A(\eta)$ as the process η conditional on $\eta \in \eta_o^A$.

$\xi^{(A)}$ is then a Poisson point process with exclusion on A.

Thus, for $M \in N$,

$$P(\xi^A(\eta) \in M) = \frac{P(\eta \in M \cap \eta_o^A)}{P(\eta \in \eta_o^A)}$$

We now take A_n compact, with $A_n \uparrow \mathbb{R}^2$, and such that $\text{int}(A_n) \uparrow \mathbb{R}^2$, and consider $\xi_n = \xi^{A_n}(\eta)$. If there is a point process ξ such that $\xi_n \xrightarrow{d} \xi$, we call ξ a Poisson point process with exclusion. We can demonstrate the existence of such a process (for suitable sets A_n) by showing that $\{\xi_n\}$ is relatively compact with respect to convergence in distribution in the vague topology.

For B bounded, let $M_t = \{\mu : \mu B > t\}$. Then, $\exists\ n_o$ s.t. $B \subseteq A_{n_o}$.

For $n \geqslant n_o$, $P(\xi_n B > t) = P(\eta \in M_t \cap \eta_o^{A_n}) / P(\eta \in \eta_o^{A_n})$

$$= \frac{1}{Z(A_n)} \sum_{k=t}^{\infty} \frac{1}{k!} \sum_{\ell=0}^{\infty} \int\cdots\int_{B^k} \int\cdots\int_{(A_1 \setminus B)^\ell} H(x_1, \ldots, x_k, z_1, \ldots, z_\ell)\, dx_1 \ldots dx_k\, dz_1 \ldots dz_\ell$$

$$\leqslant \frac{1}{t!} \int\cdots\int_{B^t} H(x_1, \ldots x_t) dx_1, \ldots dx_t \qquad (4.4)$$

$$\to 0 \quad \text{as} \quad t \to \infty;$$

here, the notation is as in Mürmann (1978), and inequality (4.4) follows as in Lemma 2.2 of that paper.

Hence, by Lemma 4.5 of Kallenberg (1975), $\{\xi_n\}$ is relatively compact. This proves the existence of an accumulation point ξ of $\{\xi_n\}$. It is straightforward to show that ξ is a Poisson process with exclusion in Mürmann's sense, in that its conditional distributions are almost surely those specified by Mürmann (1978). This is done as follows.

Let A be a compact subset of $\mathbb{R}^2$. Let $N(A)$ denote the σ-algebra in $\mathcal{N}^*$ generated by the mappings $\mu \to 1_A \mu B$, $B \in \mathcal{B}^2$, and let $M \in N(A)$. Let ξ be an accumulation point of $\{\xi_n\}$, as defined above. We consider the conditional expectation $E(I(\xi \in M) \,|\, 1_{A^c}\xi)$. Note that, because of the conditions on A_n, $P(\xi \in \mathcal{N}_o) = 1$. Let $N \in N^*$ where N^* is the σ-algebra generated by the mappings $\mu \to \mu B$, $B \in \mathcal{B}^2$, $\mu \in \mathcal{N}^*$. Then,

$$E(E(I(\xi \in M) \,|\, 1_{A^c}\xi)\ I(1_{A^c}\xi \in N)) = E(I(\xi \in M)\ I(1_{A^c}\xi \in N))$$

$$= E(I(\xi \in M)\ I(1_{A^c}\xi \in N)\ I(\xi \in \mathcal{N}_o))$$

$$= E(I(\xi \in M \cap \mathcal{N}_o)\ I(1_{A^c}\xi \in N \cap \mathcal{N}_o))$$

$$= P(\xi \in M \cap \mathcal{N}_o,\ 1_{A^c}\xi \in N \cap \mathcal{N}_o)$$

$$= \lim_{j\to\infty} P(\xi_j \in M \cap \mathcal{N}_o,\ 1_{A^c}\xi_j \in N \cap \mathcal{N}_o)$$

$$= \lim_{j\to\infty} P(\eta \in M \cap \mathcal{N}_o,\ 1_{A^c}\eta \in N \cap \mathcal{N}_o,\ \eta \in \mathcal{N}_o^{A_j}) \,/\, P(\eta \in \mathcal{N}_o^{A_j})$$

Now, for j sufficiently large, $I(A) \subseteq A_j$, so that

$$\{\eta : \eta \in M \cap \eta_o,\ 1_{A^c}\eta \in N \cap \eta_o,\ \eta \in \eta_o^{A_j}\} =$$

$$\{\eta : \eta \in M \cap \eta_o^{I(A)},\ 1_{A^c}\eta \in N \cap \eta_o\}\ .$$

Thus

$$E(E(I(\xi \in M)\,|\,1_{A^c}\xi)I(1_{A^c}\xi \in N)) = \frac{P(\eta \in M \cap \eta_o^{I(A)},\ 1_{A^c}\eta \in N \cap \eta_o)}{P(\eta \in \eta_o^{A})} \qquad (4.5)$$

From the conditional probabilities given in section 1 of Mürmann (1978), it follows that the probability on the right hand side of (4.5) is the same as that which would have been obtained using Mürmann's definition of a Poisson process with exclusion, rather than our definition of ξ. Thus the conditional distributions of ξ are almost surely the same as those for Mürmann's process.

Whether such an accumulation point ξ is unique remains an open question. To prove uniqueness, we need to show that the accumulation point is independent of the subsequence chosen, i.e. that if $\xi_{n_j} \xrightarrow{d} \xi$ and $\xi_{m_k} \xrightarrow{d} \zeta$, where $\xi_{n_j} = \xi^{A_j}(\eta)$; A_j compact, int $(A_j) \nearrow \mathbb{R}^2$, $\xi_{m_k} = \xi^{C_k}(\eta)$, C_k compact, int $(C_k) \nearrow \mathbb{R}^2$, then $\xi \overset{d}{=} \zeta$.

By 3.3 of Kallenberg (1975), it is sufficient to prove that

$$P(\xi U = 0) = P(\zeta U = 0)$$

for all bounded Borel sets U.

Let U be such a set; then I(U) is, by assumption, bounded. Thus, $\exists\ j_o$ s.t. $I(U) \subseteq A_{j_o}$ and $\exists\ k_o$ s.t. $I(U) \subseteq C_{k_o}$.

Then

$$P(\xi U = 0) = \lim_{j\to\infty} P\left(\eta U = 0,\ \eta \in \eta_o^{A_j}\right)/P\left(\eta \in \eta_o^{A_j}\right)$$

$$= \lim_{\substack{j\to\infty \\ j>j_o}} \left\{ \frac{P(\eta U = 0)\, P\left(1_{U^c}\eta \in \eta_o^{A_j \setminus U}\right)}{P\left(\eta \in \eta_o^{A_j} \,\middle|\, 1_{U^c}\eta \in \eta_o^{A_j \setminus U}\right) P\left(1_{U^c}\eta \in \eta_o^{A_j \setminus U}\right)} \right.$$

$$= \lim_{j\to\infty} P(\eta U = 0)/P\left(\eta \in \eta_o^{A_j} \,|\, 1_{U^c}\eta \in \eta_o^{A_j \setminus U}\right)$$

and similarly,

$$P(\zeta U = 0) = \lim_{k\to\infty} P(\eta U = 0)/P\left(\eta \in \eta_o^{C_k} \,|\, 1_{U^c}\eta \in \eta_o^{C_k \setminus U}\right).$$

Thus, it is sufficient to prove that

$$\lim_{j\to\infty} P\left(\eta \in \eta_o^{A_j} \,|\, 1_{U^c}\eta \in \eta_o^{A_j \setminus U}\right) = \lim_{k\to\infty} P\left(\eta \in \eta_o^{C_k} \,|\, 1_{U^c}\eta \in \eta_o^{C_k \setminus U}\right)$$

Now,

$$P\left(\eta \in \eta_o^{A_j} \,|\, 1_{U^c}\eta \in \eta_o^{A_j \setminus U}\right) = P\left(\eta \in \eta_o^{I(U)} \,|\, 1_{U^c}\eta \in \eta_o^{A_j \setminus U}\right) \ ; \text{a}$$

similar identity holds with A_j replaced by C_k.

Thus, to prove uniqueness, it is sufficient to prove, in Mürmann's notation, that as $A_j \uparrow \mathbb{R}^2$, $Z(A_j)/Z(A_j \setminus U)$ tends to a limit independent of the choice of the A_j's.

Whether this is true for I(x) as in (4.1) remains unresolved.

4. Some Properties of the Matérn 1 and 2 Models

The Matérn 1 model has intensity $\lambda \exp(-\lambda\pi R^2)$; thus two different original intensities of the Poisson process can, for a given value of R, give rise to the same final intensity. Some examples of this phenomenon are given in Table 1. Thus, the value of R and the final intensity do not uniquely determine the process for a Matérn 1 model, since processes arising from different original intensities have different properties; that this is the case can be seen by considering the second moment density g(t), given by

$$g(t) = \begin{cases} 0 & t \leq R \\ \lambda^2 \exp\{-\lambda\, U(t)\} & t > R \end{cases}$$

where

$$U(t) = m\,\{Z_R(Q) \cup Z_R((t,0))\}$$

and

$$Z_R(x) = \{y \in \mathbb{R}^2 : |x-y| < R\} \quad .$$

Figure 1 shows a plot of g(t) for two different values of λ giving rise to the same final intensity.

Although the processes have different theoretical properties, it seems difficult to distinguish between them in practice. A simple simulation experiment was carried out to distinguish between two Matérn 1 processes with the same final intensity - processes with initial intensities $\lambda = 86.089$ and $\lambda = 180.0$, and $R = 0.05$ in the unit square. The simple statistic $\hat{K}(2R)$ was used as a test statistic; $|K_{(180.0)}(x) - K_{(86.089)}(x)|$ reaches a maximum at $x = 2R$, and is constant from then on. In 100 simulations, a 5% Monte Carlo test based on $\hat{K}(2R)$ of the null hypothesis H_o: $\lambda = 180.0$ had power 13% against

TABLE 1

Some values of the original intensity, λ, which, for $R = 0.05$, give the same final intensity for the Matérn 1 model.

R	Values of λ	Final Intensity
0.05	170.0 or 92.468	44.729
0.05	180.0 or 86.089	43.783
0.05	190.0 or 80.324	42.757
0.05	200.0 or 74.923	41.609
0.05	210.0 or 69.940	40.391

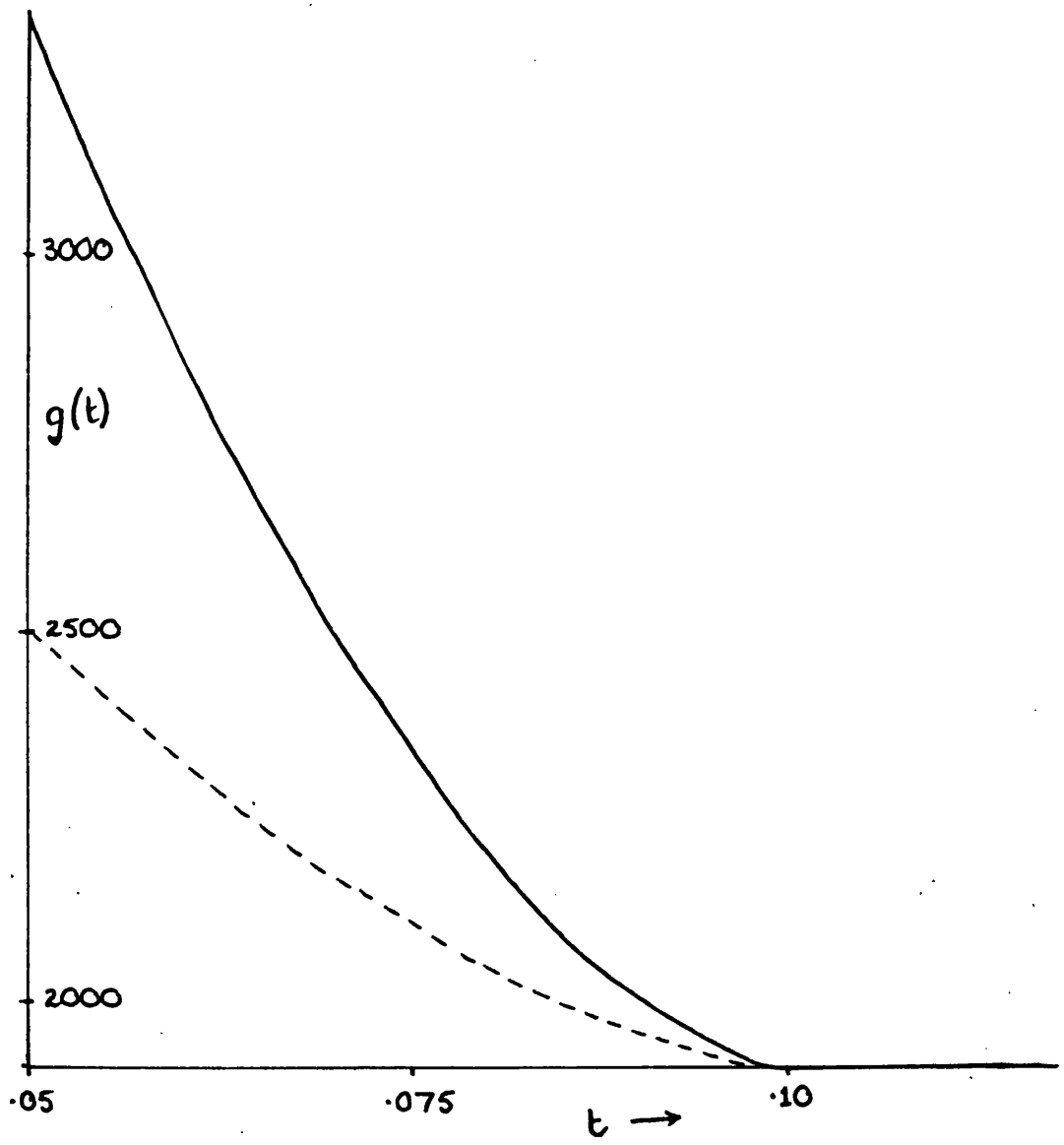


Figure 1 Graph of g(t) for two Matérn 1 models having the same disc diameter and the same final intensity. The unbroken line is g(t) for R = 0.05, λ = 180.0, while the dashed line corresponds to R = 0.05, λ = 86.089.

the alternative hypothesis H_1: $\lambda = 86.089$.

The Matérn 2 model has final intensity $\dfrac{1 - e^{-\lambda\pi R^2}}{\pi R^2}$, and its second moment density g(t) is given by

$$g(t) = \begin{cases} 0 & t \leq R \\ \dfrac{2\,U(t)(1 - e^{-\lambda\pi R^2}) - 2\pi R^2\,(1 - e^{-\lambda U(t)})}{\pi R^2 U(t)(U(t) - \pi R^2)} & t > R \end{cases}$$

- see Paloheimo (1971) for details.

Figure 2 shows g(t) plotted for Matérn 1 and Matérn 2 models with the same value of R and the same final intensity. A similar simulation test to that described above was used to attempt to distinguish between a Matérn 1 process and a Matérn 2 process with corresponding final intensities (Matérn 1, $\lambda = 180.0$; Matérn 2, $\lambda = 53.56$; $R = 0.05$ in each case). $\hat{K}(2R)$ was again used as a test statistic. In 100 simulations, a 5% Monte Carlo test based on this statistic of the null hypothesis

$$H_0'\text{: Matérn 1, } \lambda = 180.0$$

had power 26% against the alternative hypothesis

$$H_1'\text{: Matérn 2, } \lambda = 53.56$$

The results of this section suggest that there is little difference between the various Matérn models giving the same final intensity; certainly, it seems difficult to distinguish between them in practice.

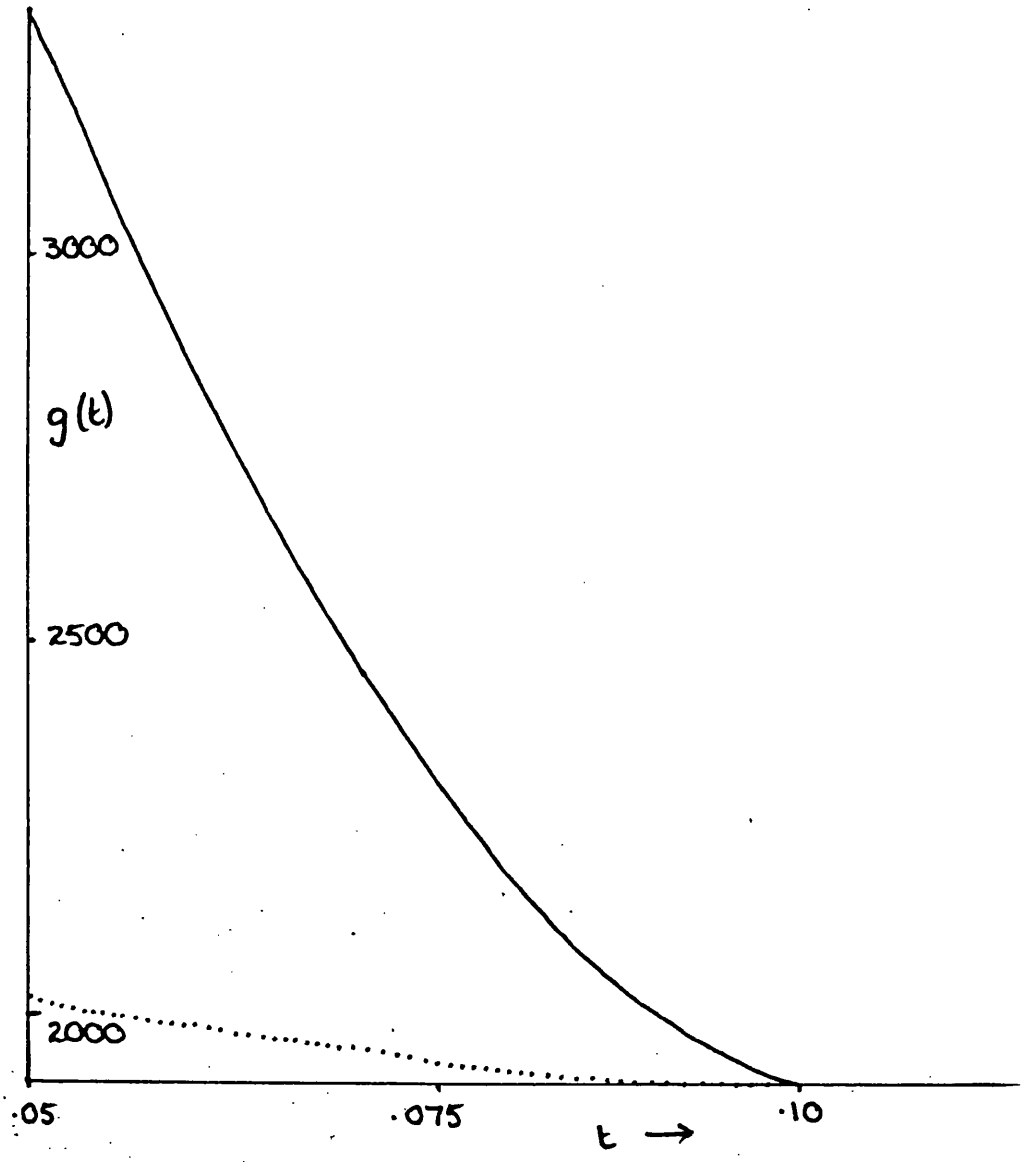


Figure 2 Graph of g(t) for a Matérn 1 and Matérn 2 model having the same disc diameter and the same final intensity. The unbroken line is g(t) for a Matern 1 model with R = 0.05, λ = 180.0, while the dotted line is g(t) for a Matérn 2 model with R = 0.05, λ = 53.654.

CHAPTER FIVE

CONVERGENCE OF SPATIAL BIRTH AND DEATH PROCESSES

1. Introduction

It is often necessary, for the purposes of testing the goodness of fit of a model, to simulate realisations of various spatial point patterns (see, for example, Besag and Diggle (1977), Ripley (1977)). Common models for spatial patterns are those described by Strauss (1975) and Kelly and Ripley (1976), where the points lie in a bounded region E, and the joint density is proportional to

$$f(x) = b^n c^s \qquad (5.1)$$

where n = number of points in the configuration

and s = number of pairs of points within distance R of each other. The case $c = 0$ is one model for a hard core process where the points represent the centres of non-overlapping discs of diameter R; this is the Kelly-Ripley model described in section 2 of Chapter 4. For simplicity, all point processes considered in this chapter will be defined on the unit square U in two-dimensional Euclidean space.

These models are difficult to simulate directly, and are most conveniently simulated using spatial birth and death processes, as described by Ripley (1977 and 1979a). In this method, the steps of a Markov process are simulated, the process being chosen so that it has as its equilibrium distribution the required Kelly-Ripley process. By allowing the process to run for a large number of steps, it should get 'near' to equilibrium, and thus the final configuration should be 'close' to what is required. Other spatial models, for example, pairwise

interaction models, are also most easily simulated using such spatial birth and death processes; see Ripley (1977 and 1979a) for more details.

In this chapter, attention is concentrated mainly on Kelly-Ripley processes with $c = 0$, i.e. on hard core processes; many of the results are then generalised to the case of other models. Coupling arguments from the theory of Markov chains are used to determine conditions under which the processes converge to a unique equilibrium distribution, and to give a lower bound on the rate of convergence. Particular use is made of 'ε-coupling', which couples two processes when they get within a distance ε of each other in a certain metric. This technique bears a certain resemblance to the coupling method introduced by Lindvall (1977). However, there is a fundamental difference: whereas Lindvall's method guarantees only that the coupled processes remain close, our 'ε-coupling', if it is successful, ensures that the two processes eventually become identical.

We shall consider two different types of spatial birth and death processes. The first of these, the general birth and death process, gives rise to processes of given intensity; the second class considered, the fixed number process, is of interest when conditioning on the total number of points observed.

2. Description of the Processes

a) The general birth and death process

These processes are special cases of spatial Markov birth and death processes as described by Preston (1976). Let x denote the state of the process at time t. In an infinitesimal interval $[t,t + \delta t)$, points are added to the process (are 'born') with spatial intensity $B_x(\xi)$, where

$$B_x(\xi) = f(x \cup \xi) / f(x) \; ; \qquad (5.2)$$

here, f(x) is the joint density of the model being simulated. For Kelly-Ripley models, where f(x) is as in (5.1), we have, for constants $b > 0$ and $c \in [0,1]$,

$$B_x(\xi) = bc^{t(\xi,x)} \; ; \qquad (5.3)$$

here $t(\xi,x)$ is the number of points of x within R of ξ. For the hard core case, $c = 0$ and so $B_x(\xi) = bI$ [ξ is not within R of any point of x].

In the interval $[t,t + \delta t)$, points are deleted from the process ('die') independently of one another, each with probability δt. For the Kelly-Ripley model, for example, it is easily shown, by Theorem 7.1 of Preston (1976), that these processes converge to a unique equilibrium process, which is the required process defined in (5.1) above.

b) The fixed number process

Starting from an initial configuration of N points, single points are alternately deleted and added. At each stage the point to be deleted is chosen uniformly from the existing points, while the point to be added is placed with intensity proportional to $B_x(\xi)$, as defined in (5.2) above, where x is the process of N-1 points remaining after the deletion.

Ripley (1979a) argues that, for the purposes of computer simulation, the statespace is finite, so that, since the Markov chain is aperiodic, convergence of the process is guaranteed provided that the chain is irreducible. It is noted by Ripley (1977) that one possible equilibrium distribution is precisely the process with joint density f(x) described above conditioned on exactly N points being present. Irreducibility of the chain would guarantee uniqueness of this equilibrium distribution.

We shall consider the convergence of the process without discretising to a finite state space, give sufficient conditions for irreducibility and investigate rates of convergence to equilibrium.

3. Convergence of the General Birth and Death Process

In this section we consider only the Kelly-Ripley process, concentrating on the hardcore case. Let $X(t)$ be a spatial birth and death process used to simulate a Kelly-Ripley model, evolving in time with transition rates as given in Section 2a, and with initial distribution λ. Let $Y(t)$ be a similar chain, started in equilibrium, i.e. having as its initial distribution the equilibrium distribution π of the process. We will investigate the convergence of the process $X(t)$ by considering the double process $Z(t) = (X(t), Y(t))$.

The state space of this double process is almost entirely continuous, but has one 'discrete' state $(0,0)$, the state where there are no points present in both $X(t)$ and $Y(t)$. Because the state space is partly discrete, the coupling argument of Pitman (1974) on discrete state space Markov chains can be modified to apply to the process $Z(t)$. Let τ be the first hit on $(0,0)$, and define

$$U(t) = \begin{cases} X(t) & t < \tau \\ Y(t) & t \geqslant \tau \end{cases} .$$

Thus we couple $X(t)$ and $Y(t)$ when they both hit 0. It follows that $U(t) \overset{\mathscr{D}}{=} X(t)$, and standard coupling arguments yield that for any measurable subset C of the space $\mathscr{B}$ of all possible configurations,

$$\left| P(X(t) \in C) - \pi(C) \right| \leqslant P(\tau > t). \qquad (5.4)$$

We thus need to investigate $P(\tau > t)$ to obtain results on the rate of convergence of $X(t)$ to equilibrium. Doing this yields the following

THEOREM A

In the case $c = 0$, there are constants γ,ρ with $0 < \gamma < \infty$ and $0 < \rho < 1$ such that, for all measurable $C \subseteq \mathcal{B}$,

$$|P(X(t) \in C) - \pi(C)| \leqslant \gamma \rho^t. \tag{5.5}$$

i.e. the convergence is geometric.

Proof

Suppose x and y are any configurations. Let $m = |x|+|y|$.

Let $\beta(x) = \int_U B_x(\xi)d\xi$ and $\beta_1 = \beta(x) + \beta(y)$.

Then, conditional on $Z(0) = (x,y)$, by standard Markov process theory, for some function $K_1(m,t_1)$, P(next event in Z occurs before time t_1 and is a death)

$$= [1 - \exp\{-(\beta_1+m)t_1\}]m/(\beta_1+m)$$

$$\geqslant K_1(m,t_1) > 0 \qquad \text{for all m, since } \beta_1 \leqslant 2b. \tag{5.6}$$

It follows that, given $Z(0) = (x,y)$, for $t_o > 0$, P(Z(t) hits (0,0) during $[0,t_o]$)

$$\geqslant P \text{ (first m events happen before } t_o \text{ and are all deaths)}$$

$$\geqslant P \text{ (first m inter-event times are at most } t_o/m \text{ and the first m events are all deaths)}$$

$$\geqslant \prod_{j=1}^{m} K_1(j,\ t_o/m) = K_o(m,t_o)$$

say, where $K_o(m,t_o) > 0$ by (5.6) above.

In the hard core case, m is bounded and so there exists $K(t_o) > 0$ such that for all possible initial configurations,

$$P(Z(t) \text{ hits } (0,0) \text{ during } [0,t_o]) \geqslant K(t_o).$$

By a similar argument to that of Doob (1953), p. 194, for any $t > t_o$,

$$P(\tau > t) \leq (1-K)^{t/t_o - 1} ;$$

putting $\gamma = (1-K)^{-1}$ and $\rho = (1-K)^{1/t_o}$ completes proof of the theorem. ∎

The inequalities developed in the first part of the proof also apply to the 'soft core' case $c > 0$, but for the moment we shall restrict attention to the hard core case.

The state 0 satisfies the requirements that the process will return to 0 infinitely often and that the mean time between successive returns is finite. Therefore, in principle at least, using the method of Crane and Iglehart (1974), the sections of the process between successive returns to zero may be used as independent and identically distributed blocks in the investigation of the equilibrium properties of the process. However, except for small packing densities, application of the Crane-Iglehart method in practice may be infeasible because the mean time between successive returns to zero, while finite, may be extremely large.

We next consider the 'fixed number' process. Since the state (0,0) is no longer accessible, the technique of coupling the two processes X and Y when they are 'close' rather than identical is used. This technique could also be applied to the general process, where it may well give better bounds on γ and ρ than those derived above; this is considered in section 8 of the chapter.

4. Distances between processes, and ε-coupling

In order to couple processes when they become 'close' it is first of all necessary to define a suitable metric on point patterns. Suppose x and y are two point patterns with the same number of points. Denote by P the set of (1-1) correspondences between the points of x and those of y. If $x = \{\xi_1, \ldots, \xi_n\}$, define

$$\rho_o(x,y) = \min_{q \in P} \max_{1 \leq i \leq n} \{d(\xi_i, q(\xi_i))\}$$

where d denotes Euclidean distance.

For general patterns x and y, let

$$\rho(x,y) = \begin{cases} \min(1, \rho_o(x,y)) & \text{If x and y have the same number of points} \\ 1 & \text{otherwise} \end{cases}$$

It follows easily that ρ is a metric on point patterns. Note that ρ(x,y) is at least as large as the minimum of 1 and the Prohorov distance p(x,y) between the atomic measures generated by x and y. (See Billingsley (1971), pages 237-238 for the definition of the Prohorov metric). In fact, it can be shown that

$$\rho(x,y) = \min(1, p(x,y)). \qquad (5.7)$$

The proof of this fact is based on the following theorem of Philip Hall (Hall, 1935):

Theorem

If $S_1, \ldots, S_N$ are distinct subsets of a finite set W, they have a system of distinct representatives (i.e. ∃ distinct elements $a_1, \ldots, a_N$ with $a_i \in S_i$, i = 1, ..., N) if and only if for each k = 1, ..., N,

any k distinct sets contain between them at least k distinct elements.

Proof of (5.7)

Let x and y be two point patterns in U. If B is a Borel subset of U, denote by $N_x(B)$ and $N_y(B)$ the number of x-points and y-points, respectively, contained in B. For $\eta > 0$, let $B^\eta = \{u \in U : \exists v \in B$ with $d(u,v) < \eta\}$. From the definition of the Prohorov metric,

$p(x,y) < \eta$ iff for all closed subsets A of U

$$N_x(A) \leqslant N_y(A^\eta) + \eta$$

and $$N_y(A) \leqslant N_x(A^\eta) + \eta \ .$$

Note that if $\eta < 1$, this reduces to

$p(x,y) < \eta$ iff for all closed subsets A of U

$$N_x(A) \leqslant N_y(A^\eta)$$

and $$N_y(A) \leqslant N_x(A^\eta) \ .$$

From this it is clear that $\rho(x,y) \geqslant \min(1, p(x,y))$. To prove that $\rho(x,y) \leqslant \min(1, p(x,y))$, proceed as follows. (I am grateful to Prof. R. Sibson for pointing out that the result could be proved by this method).

Suppose that $0 < \eta < 1$, and that $p(x,y) < \eta$. Let $x = \{\xi_1, \ldots, \xi_N\}$. Since $\eta < 1$, y must also contain N points.

Let $S_i = \{u \in y : |u - \xi_i| < \eta\}$ $\quad i = 1, \ldots, N.$

Since $p(x,y) < \eta$, we have that, if $A = \{\xi_{i_1}, \ldots, \xi_{i_k}\}$, with $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant N$, then

$$|S_{i_1} \cup \ldots \cup S_{i_k}| = N_y(A^\eta) \geqslant N_x(A) = k \ .$$

Thus, for each $k = 1, \ldots, N$, any k distinct sets from the S_i's

contain between them at least k distinct elements.

Hence, by P. Hall's theorem, there exist distinct elements $a_1, \ldots, a_N$ with $a_i \in S_i$, $i = 1, \ldots, N$. Defining $q(\xi_i) = a_i$, $i = 1, \ldots, N$, we have a (1-1) correspondence q between the points of x and those of y, with corresponding points distance less than η apart. Thus $\rho(x,y) < \eta$, completing the proof of (5.7). ∎

Another possible metric for point patterns is described by Matthes, Kerstan and Mecke (1978, p. 108); our metric ρ gives a natural criterion for the closeness of two point patterns which suits our purposes.

Once the distance ρ between point patterns has been defined, it is possible to define a distance d_ρ between point processes, viewed as random point patterns; given point processes X and Y, we define

$$d_\rho(X,Y) = \inf \{\eta > 0 : P(\rho(X(\omega), Y(\omega)) > \eta) < \eta\} .$$

Analogously to Exercises 8 and 9 of page 94 of Chung (1974), if X and Y are close in the metric d_ρ then their 'distributions' will be close in the obvious analogy of the Lévy distance on distributions over the metric space of point patterns with metric ρ.

To define the coupling, given a positive number $\varepsilon < 1$ and two birth and death processes $\{X_n\}$ and $\{Y_n\}$, we wait until $\rho(X_n,Y_n) < \varepsilon$. A (1-1) correspondence can then be set up between the points of the two configurations X_n and Y_n, with points corresponding to each other being a distance at most ε apart. The two processes are then coupled as follows:

i) When deaths occur, corresponding points are deleted.

ii) With a certain (high) probability (depending on how close together the patterns are), the same point is born for each process;

otherwise different points are inserted for the two processes, in which case the coupling fails and the processes subsequently evolve independently.

The coupling will be successful provided we do not insert different points for the two processes before all the points present when the coupling was started are deleted. By making ε small, the probability that the coupling is successful can be made as close to 1 as we wish. Details of the coupling are given in full in the next section.

5. Proof of convergence of the fixed number process, using ε-coupling

Let $\{X_n\}$ and $\{Y_n\}$ be two realisations of the fixed number birth and death process with N points, as described in section 2(b) above. Each is observed after each birth, so that for each process, one step consists of a death followed by a birth. Let $\{X_n\}$ have arbitrary initial distribution λ, with $\{Y_n\}$ having initial distribution π_N, the equilibrium distribution of the process. Thus the chain $\{Y_n\}$ is in equilibrium. We will couple $\{X_n\}$ and $\{Y_n\}$ as described in section 4 above. We first need some notation. For any (N-1)-point pattern x, and $\xi \in U$, define

$$C_x(\xi) = B_x(\xi) / \int_{u \in U} B_x(u)\,du$$

where $B_x(\xi)$ is as defined in (5.2) above. For any two (N-1) point patterns x and y, and $\xi \in U$, define

$$g_{x,y}(\xi) = \min\,(C_x(\xi),\ C_y(\xi))$$

$$G_{x,y} = \int_{\xi \in U} g_{x,y}(\xi)\,d\xi.$$

To define the coupling formally, let

$$T_\varepsilon = \min\,\{n : \rho(X_n, Y_n) < \varepsilon\}\ .$$

Define a new process $Z_n^{(\varepsilon)} = (V_n, W_n)$ where

i) $V_n = X_n$ and $W_n = Y_n$ for $n \leq T_\varepsilon$

ii) When $n = T_\varepsilon$, let σ be the (1-1) correspondence for which the minimum in the definition of $\rho(X_n, Y_n)$ is attained. The processes then evolve as follows. First kill a point of each process by setting $(x,y) = (V_n \setminus \xi,\ W_n \setminus \sigma(\xi))$ where ξ is chosen uniformly from V_n. The

'birth' part of the step is a little more complicated.

With probability $G_{x,y}$, choose ζ in U according to the probability density $g_{x,y}(\zeta)/G_{x,y}$, define $\sigma(\zeta) = \zeta$, set $(V_{n+1}, W_{n+1}) = (x \cup \zeta, y \cup \zeta)$ and continue the attempt to couple the processes. (Thus we have allowed the same point to be born for each process). Otherwise abandon the attempt to couple the processes and in subsequent steps let the processes evolve independently. To complete this step set $(V_{n+1}, W_{n+1}) = (x \cup \zeta_1, y \cup \zeta_2)$ where ζ_1 is chosen in U with density proportional to $C_x(\zeta_1) - g_{x,y}(\zeta_1)$ and ζ_2 is chosen in U with density proportional to $C_y(\zeta_2) - g_{x,y}(\zeta_2)$.

It is easily verified that V_n and W_n will individually have the same probability structure as X_n and Y_n respectively. Note that if all the points present at T_ε have died before the coupling fails, then the processes will remain identical; in this event call the coupling "successful". We use ε-coupling to prove the following result:

THEOREM B

Define X_n and Y_n as above, and define d_ρ as in section 4. Let $D_n = (X_n, Y_n)$ and $A_\varepsilon = \{(x,y) : \rho(x,y) < \varepsilon\}$ and define $E(x,y) = \int_{u \in U} |C_x(u) - C_y(u)| du$. Suppose that

i) $E(x,y) = O(\rho(x,y))$ as $\rho(x,y) \to 0$

and

ii) there exist n_o and $K_\varepsilon > 0$, with $K_\varepsilon = O(\varepsilon^{2N})$ such that for all possible configurations x and y,

$P(D_n$ enters A_ε during the first n_o steps $\mid D_o = (x,y)) \geq K_\varepsilon$.

If $\eta > 0$ then, on a suitable probability space, there exist versions V_n and W_n of X_n and Y_n respectively such that

$$d_\rho(V_n, W_n) = o(n^{-\frac{1}{2N} + \eta})$$

as n tends to infinity.

It will then follow, as noted in section 4 above, that the "ρ-Levy" distance between X_n and Y_n is $o(n^{-\frac{1}{2N} + \eta})$ as n tends to infinity, for all $\eta > 0$.

Before proving the theorem, we need to prove three lemmas.

Lemma 1

Suppose there exist n_o and $K_\varepsilon > 0$ such that for all possible N-point configurations x and y,

$P(D_n$ enters A_ε during the first n_o steps $\mid D_o = (x,y)) \geqslant K_\varepsilon$. Then there exist constants α and δ $0 < \delta < 1$ and $0 < \alpha < \infty$ such that

$$P(T_\varepsilon > n) \leqslant \alpha\delta^n, \tag{5.8}$$

i.e. T_ε has a geometric tail.

Proof

Note that $T_\varepsilon = \min \{n: D_n \epsilon A_\varepsilon\}$. Using an argument similar to that of Doob (1953), page 194, we have that

$$P(T_\varepsilon > n) \leqslant (1 - K_\varepsilon)^{\frac{n}{n_o} - 1}. \tag{5.9}$$

Setting $\alpha = (1 - K_\varepsilon)^{-1}$ and $\delta = (1 - K_\varepsilon)^{\frac{1}{n_o}}$ proves the lemma. ∎

Lemma 2

Suppose that condition (i) of Theorem B holds. Then

$$P(\varepsilon\text{-coupling successful}) \to 1 \text{ as } \varepsilon \to 0 \tag{5.10}$$

Proof

$\frac{1}{\lambda}E(x,y)$ is the probability that, at a stage in the attempted coupling when the current (N-1)-point patterns are x and y, the coupling fails.

Since each step in an attempted coupling does not increase the ρ-distance between the patterns, it follows easily from condition (i) and the remark immediately before Theorem B that

$$P(\varepsilon\text{-coupling successful}) \geqslant \left(\frac{p}{p+p_\varepsilon}\right)^N$$

where p is a constant and $p_\varepsilon = O(\varepsilon)$ as $\varepsilon \to 0$.

For $C \subseteq \mathscr{B}$ and $\eta > 0$, define

$$C^\eta = \{x \in \mathscr{B} : \exists\ y \in C \text{ with } \rho(x,y) < \eta\} .$$

Lemma 3

Under the conditions of Theorem B, the process X_n converges to a unique equilibrium distribution π_N in the sense that, given $\eta > 0$, there is an n_1 such that, for all measurable $C \subseteq \mathscr{B}$,

$$\text{and} \qquad \begin{aligned} P(X_n \in C) &\leqslant \pi_N(C^\eta) + \eta \\ \pi_N(C) &\leqslant P(X_n \in C^\eta) + \eta \end{aligned} \qquad \text{for } n \geqslant n_1 .$$

Proof

By (5.10), we may choose $\varepsilon < \eta$ such that $P(\varepsilon\text{-coupling is unsuccessful}) < \eta/2$.

By (5.8), we can choose an n_1 such that

$$P(T_\varepsilon > n) \leqslant \eta/2 \qquad \text{for all } n \geqslant n_1 .$$

The construction of $V_n^{(\varepsilon)}$ and $W_n^{(\varepsilon)}$ is such that, for all measurable $C \subseteq \mathscr{B}$,

$$\text{and} \qquad \begin{aligned} P(X_n \in C) &\leqslant P(Y_n \in C^\varepsilon) + p_{n,\varepsilon} \\ P(Y_n \in C) &\leqslant P(X_n \in C^\varepsilon) + p_{n,\varepsilon} \end{aligned}$$

where $p_{n,\varepsilon} = 1 - P(T_\varepsilon \leqslant n \text{ and coupling successful})$.

Since $P(Y_n \in C) = \pi_N(C)$, and

$P_{n,\varepsilon} \leqslant P(T_\varepsilon > n) + P(\varepsilon\text{-coupling unsuccessful})$, the result now follows from the above and the fact that $C^\varepsilon \subseteq C^\eta$. ∎

We can now prove Theorem B.

Proof of Theorem B

Define V_n and W_n as above.

From ε-coupling we know that

$$P(\rho\{V_n(\omega), W_n(\omega)\} > \varepsilon) \leqslant P(T_\varepsilon > n) + P(\varepsilon\text{-coupling unsuccessful})$$

$$= \gamma_\varepsilon(n) + \beta_\varepsilon, \quad \text{say.}$$

Now, $\beta_\varepsilon \leqslant 1 - \left(\frac{p}{p+p_\varepsilon}\right)^N$, where $p_\varepsilon = O(\varepsilon)$, so there exists $k_1 > 0$ such that $\beta_\varepsilon \leqslant k_1 \varepsilon$.

From the bounds on $P(T_\varepsilon > n)$,

$$\gamma_\varepsilon(n) \leqslant \alpha \delta_\varepsilon^n$$

where $\delta_\varepsilon = (1 - K_\varepsilon)^{1/n_o}$ and, in terms of ε and N, $K_\varepsilon \geqslant k_2 \varepsilon^{2N}$ for some $k_2 > 0$. Thus

$$\delta_\varepsilon^n \leqslant (1 - k_2 \varepsilon^{2N})^{n/n_o}$$

and so there exists $k_3 > 0$ such that

$$\delta_\varepsilon^n \leqslant \exp(-k_3 n \varepsilon^{2N}).$$

Thus there exists $k_4 > 0$ such that

$$P(\rho(V_n(\omega), W_n(\omega)) > \varepsilon) < k_4 \exp(-k_3 n \varepsilon^{2N}) + k_1 \varepsilon$$

$$= u(\varepsilon), \quad \text{say.}$$

Since $\alpha = (1 - K_\varepsilon)^{-1}$ and $K_\varepsilon = O(\varepsilon^{2N})$, k_4 can be chosen to be independent of ε.

It follows immediately that

$$P(\rho(V_n(\omega), W_n(\omega)) > \varepsilon) \leqslant \inf_{\varepsilon^* \leqslant \varepsilon} \{u(\varepsilon^*)\} = \tilde{u}(\varepsilon) \text{ say.}$$

By definition, $\tilde{u}(\varepsilon)$ is continous and non-increasing. Let $\varepsilon_o(n)$ be the unique value such that $\tilde{u}(\varepsilon_o) = \varepsilon_o$.

Then it follows at once that

$$d_\rho(V_n, W_n) \leqslant \varepsilon_o(n).$$

Given $\eta > 0$, we now show that $\varepsilon_o = o\left(n^{-\frac{1}{2N} + \eta}\right)$ to complete the proof.

Given A, set $\varepsilon_n = An^{-\frac{1}{2N} + \eta}$ and $\varepsilon_n^* = n^{-\frac{1}{2N} + \eta/2}$. Then, for sufficiently large n, $\varepsilon_n^* < \varepsilon_n$ and so, by the definitions of $\tilde{u}$ and u,

$$\begin{aligned}\tilde{u}(\varepsilon_n) &\leqslant u(\varepsilon_n^*) \\ &= k_4 \exp(-k_3 n^{N\eta}) + k_1 n^{-\frac{1}{2N} + \eta/2} \\ &< \varepsilon_n.\end{aligned}$$

for sufficiently large n. It follows that, for sufficiently large n, $\varepsilon_o < \varepsilon_n$, by the definition of ε_o. Thus $\varepsilon_o = o\left(n^{-\frac{1}{2N} + \eta}\right)$, completing the proof of the theorem. ∎

Having proved this general theorem, we can now use it to prove, using ε-coupling, the convergence to equilibrium of the hardcore birth and death process under certain conditions on the packing density, the general Strauss birth and death process and other similar processes. This is done in the next section.

6. Proofs of convergence for particular models

a) The hardcore case

We can now prove the convergence of the hardcore process. To do this, we need to impose a condition on the packing density of the configuration.

Lemma 4

Let X_n be a hardcore birth and death process as described in section 2b).

Suppose that $N < 2(1 + R/2)(1 + \sqrt{3}\,R)/(3\sqrt{3}R^2)$. (5.11)

Then the process X_n converges to a unique equilibrium distribution π_N in the sense of Lemma 3.

Note The condition on N ensures that, given any hard core (radius $\frac{1}{2}$ R) configuration of N points, we can always add a point somewhere and still satisfy the hard core condition.

Proof of Lemma 4

To prove the lemma, we demonstrate the existence of K_ε and n_o as in Lemma 1. The result then follows from Lemma 3.

Let $A(M,R) = \inf_x\{m(A_x) : |x| = M \text{ and } x \text{ is hard core } (R)\}$.

and $B(M,R) = \sup_x\{m(A_x) : |x| = M \text{ and } x \text{ is hard core } (R)\}$.

The condition on N ensures that $\exists\ \varepsilon > 0$ such that $A(N,R + 2\varepsilon) > 0$. We assume that ε has been chosen so that this condition holds, and $A(N - 1, R + 2\varepsilon) > 4\varepsilon(1 - \varepsilon)$. We also assume that $\varepsilon < 0.3R$. For any configuration z, let $d_{min}(z) = \min\{d(\xi,\eta) : \xi,\eta \in z\}$.

We establish the existence of K_ε and n_o by showing that for any two N-point configurations x and y, transition is possible from (x,y) into

A_ε in 6N-15 steps. This is done by allowing x to evolve into any configuration z, provided

$$d_{min}(z) > R+2\varepsilon \text{ and } z \text{ has no points within } \varepsilon \text{ of the edge of } U, \quad (5.12)$$

and by ensuring that y evolves into a configuration that is within ε of z. We have

$$\begin{aligned} &P(D_n \text{ hits } A_\varepsilon \text{ in the first 6N-15 steps } | D_o = (x,y)) \\ &> P^{(6N-15)}(x \to \text{some } z \text{ satisfying condition 5.12}) \\ &\times P^{(6N-15)}(y \to N_\varepsilon^z \,|\, x \to z) \end{aligned} \quad (5.13)$$

where $N_\varepsilon^z = \{u : \rho(u,z) < \varepsilon\}$.

We consider the two expressions on the right-hand side of (5.13) separately, looking firstly at the second term. Consider any two configurations y and z, with z satisfying condition (5.12). For any point $\xi \in z$, and $\varepsilon < 0.3R$, there are at most six y-discs which overlap the circle of radius $R + \varepsilon$ centred at ξ. To move from y to N_z^ε in 6N-15 or fewer steps, remove each of these discs in turn, putting it down elsewhere, and put the final disc down so that its centre lies in the circle of radius ε centred at ξ. Repeat the process for each of the points of z. For the last point η of z, at most one disc needs to be dealt with; for the penultimate point, at most two discs, etc. Thus, in at most 6N-15 steps, Y_n can move by this procedure from y to a configuration which lies in N_z^ε.

There are N! orders in which we can deal with the points of z, and for each of the first N-6 of these points, there are 6! orders in which to remove the discs near that point.

It follows that

$$p^{(6N-15)}(y \to N_z^{\varepsilon}) \geq \left\{\frac{\pi\varepsilon^2}{B(N-1,R)}\right\}^N N! \left(\frac{6!}{N^6}\left\{\frac{A(N,R+\varepsilon)}{B(N-1,R)}\right\}^5\right)^{N-6}$$

$$\times \prod_{r=1}^{5} \frac{r!}{N^r}\left\{\frac{A(N,R+\varepsilon)}{B(N-1,R)}\right\}^{r-1}$$

$$= C(N,R,\varepsilon) \text{ say, where } C(N,R,\varepsilon) > 0.$$

We now consider the first term on the right-hand side of (5.13). We wish to move in 6N-15 steps from x to a configuration z satisfying condition (5.12). We do this by ensuring that all the original x-points die by the $(6N-15)^{th}$ step and that the configuration remaining after 6N-15 steps satisfies condition (5.12). To do this, we must have at least N points being born at least $R + 2\varepsilon$ from each existing point and at least ε from the edge of U. We can afford to make 'mistakes' (i.e. allow a point to be born too close to an existing point or to the edge) provided each such point dies by the $(6N-15)^{th}$ step.

Letting r be the number of mistakes made, some manipulation yields that the number of different orders in which the 6N-15 operations can be performed is at least

$$\binom{\alpha}{r} - \binom{\alpha-2}{r-2}\left\{(5N-15)^3 - 3(5N-15)^2 + 2(5N-15)\right\}/3$$

$$+ \binom{\alpha-4}{r-4}\left\{(5N-15)^3 - 3(5N-15)^2 + 2(5N-15)\right\}^2/12 \quad = f(N,r), \text{ say}$$

where $\alpha = \alpha(N) = (5N-15)(5N-16)/2$.

Now, the area available to put points down in 'permitted' positions is at least $A(N-1,R+2\varepsilon) - 4\varepsilon(1-\varepsilon)$. Thus the probability that a birth is 'permissible' is always at least

$$\{A(N-1,R+2\varepsilon) - 4\varepsilon(1-\varepsilon)\}/B(N-1,R) = X, \text{ say.}$$

Since each 'mistake' is a birth in the excluded region, followed at some stage by the death of that point, it follows that

$$P^{(6N-15)}(x \to \text{some } z \text{ satisfying condition (5.12)})$$

$$\geq \binom{6N-15}{N} \frac{N!}{N^N} \sum_{r=0}^{5N-15} f(N,r)\, X^{6N-15-r} Y^r$$

$$= P(N,R,\varepsilon), \text{ say,}$$

where $Y = (1-X)/N$. Now, $P(N,R,\varepsilon) > 0$. Setting $n_o = 6N-15$ and $K_\varepsilon = C(N,R,\varepsilon) \times P(N,R,\varepsilon)$ completes the proof of the lemma. ∎

THEOREM C (Rate of convergence for hard core process)

Let $\{X_n\}$ and $\{Y_n\}$ be two realisations of the fixed number hard core birth and death process with N points, $\{X_n\}$ having arbitrary initial distribution λ, and $\{Y_n\}$ having intial distribution π_N, the equilibrium distribution of the process.

If $N < 2(1 + R/2)(1 + \sqrt{3}R)/(3\sqrt{3}R^2)$, and if $\eta > 0$, then, on a suitable probability space, there exist versions V_n and W_n of X_n and Y_n respectively such that

$$d_\rho(V_n, W_n) = o(n^{-\frac{1}{2N} + \eta})$$

as n tends to infinity.

Proof

We need only check that conditions i) and ii) of Theorem B hold. Condition i) is easily checked (note that $E(x,y)$ is simply the area of the difference between the 'permitted regions' of the two patterns x and y), and condition ii) follows from Lemma 4. ∎

b) The soft core case

Theorem B can be applied to prove convergence of the soft core birth and death process; in this case, no restriction is required on the number of points present.

THEOREM D (Rate of convergence for soft core process)

Let $\{X_n\}$ and $\{Y_n\}$ be two realisations of a fixed number soft core birth and death process as described in section 2, with N points, $\{X_n\}$ having arbitrary initial distribution λ and $\{Y_n\}$ having initial distribution π_N, the equilibrium distribution of the process.

If $\eta > 0$, then, on a suitable probability space, there exist versions V_n and W_n of X_n and Y_n respectively such that

$$d_\rho(V_n, W_n) = o(n^{-\frac{1}{2N} + \eta})$$

as n tends to infinity.

Proof

Again, we need only check that conditions i) and ii) of Theorem B hold. It is easily shown that condition i) holds. To demonstrate that condition ii) holds, set $n_o = N$; for any two configurations x and y, we have, analogously to (5.13), using the same notation as in section 6 a),

$$P(D_n \text{ hits } A_\varepsilon \text{ in the first N steps}) \geq P^{(N)}(x \to \text{some z satisfying condition (5.12)}) \times P^{(N)}(y \to N_\varepsilon{}^z \mid x \to z)$$

Arguments similar to those of section 6 a) yield

$$P^{(N)}(y \to N_\varepsilon{}^z \mid x \to z) \geq \frac{(N!)^2}{N^N} \left[\frac{\pi\varepsilon^2 c^N}{1 - \pi R^2(1 - c^N)}\right]^N$$

and

$$P^{(N)}(x \to z \text{ satisfying condition (5.12)}) \geq \frac{N!}{N^N}\left[\frac{(1-N\pi\varepsilon^2 - (1-\varepsilon)\varepsilon)c^N}{1 - \pi R^2(1 - c^N)}\right]^N$$

so that condition ii) holds, completing the proof. ∎

c) Other models

Theorem B can also be used to obtain convergence results for birth and death processes associated with other spatial models; for example, more general pairwise interaction processes (Ripley, 1977, section 3.4). Here,

$$f(x) \propto \prod_{\substack{\xi,\eta \in x \\ \xi \neq \eta}} h\{d(\xi,\eta)\}$$

for some interaction function $h: (0,\infty) \to [0,\infty)$. Again, conditions i) and ii) need to be checked - this will depend on the particular form of the interaction function h. Sufficient conditions for condition i) to hold are that $h(x)$ is bounded and continuous for $x \in (0,a]$ for some $a > 0$, and that $h(x) = 0(x)$ as $x \to 0$.

7. Expected Waiting Time to Coupling

Having used ε-coupling to prove the convergence of the hard core process, it is of interest to investigate how long we need to wait until the two processes get within ε of each other and the coupling starts.

Interest centres on the behaviour of K_ε as $N \to \infty$. If we allow N to tend to infinity, we must also have R (and ε) decreasing. We let $\varepsilon = \mu R$ for some constant μ. If we assume that we can vary R with N in such a way as to allow us to treat X, A(N-1, R+ε) and B(N-1,R) as constant, and satisfy inequality (5.11), it can be shown that, for sufficiently large N,

$$K_\varepsilon(N,R) \geq \frac{HG^N\varepsilon^{2N}}{N^{5(N-5)}}$$

where H and G are constants. Using the right-hand side of (5.9) as an approximation to $P(T_\varepsilon > n)$ and comparing with the geometric distribution with corresponding tail probabilities $\delta^n(1-\delta)$ gives the following upper bound on the expected waiting time until the two processes X_n and Y_n get within ε of each other:

$$\text{Expected waiting time} \doteq (6N-15)/K_\varepsilon.$$

As $N \to \infty$, this varies approximately as

$$\frac{N^{5N-24}}{BG^N\varepsilon^{2N}}$$

where B and G are constants.

For the soft core case $c > 0$, the corresponding bound for the expected waiting time to coupling varies approximately as

$$\frac{A^N c^{-2N^2}}{N^{N+\frac{1}{2}}\varepsilon^{2N}}$$

as $N \to \infty$, where A is a constant.

8. Extensions

In this section some details are given of various possible extensions and generalisations of the coupling methods described in the preceding sections.

i) The use of ε-coupling for the general birth and death process

As was noted in section 4, ε-coupling can be used to give different bounds on the rate of convergence to equilibrium; some details are given below.

Let $X(t)$, $Y(t)$ be as in section 4, and let $\varepsilon > 0$. Let

$$\tau_\varepsilon = \inf\{t : \rho(X(t), Y(t)) < \varepsilon\} .$$

Define a new process $Z^{(\varepsilon)}(t) = (V(t), W(t))$ where

i) $V(t) = X(t)$ and $W(t) = Y(t)$ for $t \leqslant \tau_\varepsilon$

ii) When $t = \tau_\varepsilon$, let σ be the (1-1) correspondence for which the minimum in the definition of $\rho(X(t), Y(t))$ is attained.

Let $B(\xi) = \min(B_{V(t)}(\xi), B_{W(t)}(\xi))$, $\xi \in U$. Then, if $V(t) = \{\xi_1, \ldots, \xi_m\}$, the processes evolve as follows: the following transitions are possible from the state $(V(t), W(t))$.

1) $(V(t), W(t)) \to (V(t) \setminus \xi_i, W(t) \setminus p(\xi_i))$ at unit rate, $i = 1, \ldots, m$.

2) $(V(t), W(t)) \to (V(t) \cup \xi, W(t) \cup \xi)$ at rate $B(\xi)$, $\xi \in U$.

3) $(V(t), W(t)) \to (V(t) \cup \xi, W(t))$ at rate $B_{V(t)}(\xi) - B(\xi)$, $\xi \in U$.

4) $(V(t), W(t)) \to (V(t), W(t) \cup \xi)$ at rate $B_{W(t)}(\xi) - B(\xi)$, $\xi \in U$.

If the next transition is of the form 1) or 2), the coupling continues, the (1-1) correspondence between the points being adjusted in the obvious way; otherwise, the coupling fails. The coupling will be successful provided all the points present at time τ_ε die before a transition of the

form 3) or 4) occurs.

For the hard core process, since the total number of points is bounded, it is easily shown that P(ε-coupling breaks down) = $O(\varepsilon)$, so that P(ε-coupling successful) → 1 as ε → 0.

To complete the use of the ε-coupling argument for the hard core case, we need to demonstrate the existence of a t_o and a $K_\varepsilon > 0$ such that

$$P\{\rho(X(t),Y(t)) < \varepsilon \text{ for some } t \in [0,t_o] \mid (X(0),Y(0)) = (x,y)\} \geq K_\varepsilon$$

for all configurations x and y.

This is easily done using devices analogous to those of section 5 (note that we can, if desired, use the same t_o as in section 3); arguments similar to those of Lemmas 1 to 4 can then be used to prove the convergence of the process and obtain a bound on the rate of convergence to equilibrium. Since

$$P(Z(t) \text{ hits } (0,0) \text{ during } [0,t_o]) < P(Z(t) \text{ hits } A_\varepsilon \text{ during } [0,t_o]),$$

K_ε can be made strictly greater than the K obtained in section 3.

ii) Other extensions

a) Deletion of points in order

For the fixed number process, Ripley (1979a) suggested always deleting the oldest surviving point instead of choosing the point to be deleted uniformly from the existing points. This device may be applied equally to the general process; in both cases the equilibrium distribution will be unaffected and it seems plausible that convergence to equilibrium will be more rapid.

The methods of proof of sections 3 and 5 are easily generalised to deal with this extension. In the case of the general process, using

coupling at zero, the bounds on the time to coupling will be identical. In the fixed number case, an adaptation of the ε-coupling argument can be applied. The distance between configurations x and y will be redefined as

$$\rho_1(x,y) = \max_{1 \leq i \leq n} d(x^{(i)}, y^{(i)})$$

where $x^{(i)}$ denotes the i^{th} oldest point in the x pattern. Putting $p(x^{(i)}) = y^{(i)}$ in the definition of ρ_o, it is immediate that ρ_o is dominated by ρ_1, and so coupling when ρ_1 is small will always take at least as long as coupling when ρ_o is small. Thus, if it is indeed the case that the modified process converges more rapidly, it will be necessary to use methods other than ε-coupling to prove this.

b) It is possible to prove convergence of the fixed number process using a coupling device similar to the steps of ε-coupling, but without having to wait until the two patterns are within distance ε of each other: pair the points of the X- and Y- patterns at the very beginning, and then couple the processes according to step ii) of section 5. For the soft core case, there is positive probability of a complete coupling in N steps regardless of the configurations. This is also true for the hard core case provided $2(N-1)\pi R^2 < (0.906)^{-1}$.

It then follows that there exist $K > 0$ and ρ with $0 < \rho < 1$ such that

$$P\{X_n \neq Y_n\} \leq K\rho^n \quad .$$

We can choose the pairing of the points so that the ρ-distance between X_o and Y_o is minimised; even so, the probability of a successful coupling will be extremely small, and the bound on the rate of convergence obtained in this way will be very slow. In addition, for the hard core

case, this argument works only for packing densities below about 0.14, since we require that $2(N-1)\pi R^2 < (0.906)^{-1}$.

c) The add-delete process

For simulating patterns with a fixed number of points, Ripley (1977, Section 5) suggests an alternative type of birth and death process to that of section 2b). Here, starting from an N-point configuration, points are alternately added and deleted. At each stage, the point to be added is chosen uniformly in the unit square, while, at the death part of the step, the point ξ_i is deleted from the pattern $x = \{\xi_1, \ldots, \xi_{N+1}\}$ with probability proportional to $f(x \setminus \xi_i)$, Ripley (1977, section 5) notes that this type of birth and death process works only if f is strictly positive; thus it cannot be used to simulate the hard core process.

For such birth and death processes, a form of ε-coupling can be constructed as follows. As before, let

$$T_\varepsilon = \min \{n: \rho(X_n, Y_n) < \varepsilon\}$$

and define $Z_n = (V_n, W_n)$ by

i) $V_n = X_n$ and $W_n = Y_n$ for $n \leq T_\varepsilon$.

ii) For $n > T_\varepsilon$, each of V_n and W_n evolves according to a birth-death sequence of steps.

For the birth part of each step, the same point ξ is born for each of V_n and W_n. This time the death part of the step is more complicated. Let $x = V_n \cup \xi$, $y = W_n \cup \xi$; let p be the minimal (1-1) correspondence between the points of x and those of y. If $x = \{\xi_1, \ldots, \xi_{N+1}\}$, write

$g(x) = \sum_{i=1}^{n} f(x \setminus \xi_i)$, and let $\beta_i = \min\{f(x \setminus \xi_i)/g(x), f(y \setminus p(\xi_i))/g(y)\}$, $i = 1, \ldots, N+1$. Then, with probability $\beta = \sum_{j=1}^{N+1} \beta_j$, allow corresponding

points to die. Given that this happens, point i dies with probability β_i/β, and the coupling continues.

Otherwise, abandon the attempt to couple the processes and in subsequent steps let the processes evolve independently. To complete this step, set $(V_{n+1},W_{n+1}) = (x \setminus \xi_i, y \setminus p(\xi_j))$, where ξ_i is deleted with probability $\dfrac{f(x \setminus \xi_i)/g(x) - \beta_i}{1 - \beta}$ and $p(\xi_j)$ is deleted with probability $\dfrac{f(y \setminus p(\xi_j))/g(y) - \beta_j}{1 - \beta}$. Again, V_n and W_n will individually have the same probability structure as X_n and Y_n respectively.

In order to ensure that

$$P(\varepsilon\text{-coupling successful}) \to 1 \quad \text{as} \quad \varepsilon \to 0, \quad \text{we}$$

need

$$\sum_{1}^{N+1} \left| f(x \setminus \xi_i)/g(x) - f(y \setminus p(\xi_i))/g(y) \right| \to 0 \qquad \text{as}$$

$$\rho(x,y) \to 0 \,. \tag{5.14}$$

For a pairwise interaction process (see section 6c)), a sufficient condition for (5.14) to hold is that $h(x)$ be uniformly continuous on $(0,\sqrt{2}]$. Note that condition (5.14) is not satisfied for soft core Kelly-Ripley processes. If (5.14) and condition ii) of Theorem B are satisfied, ε-coupling can be used to prove convergence to equilibrium of the add-delete process, in a way analogous to that of section 5.

9. Conclusion

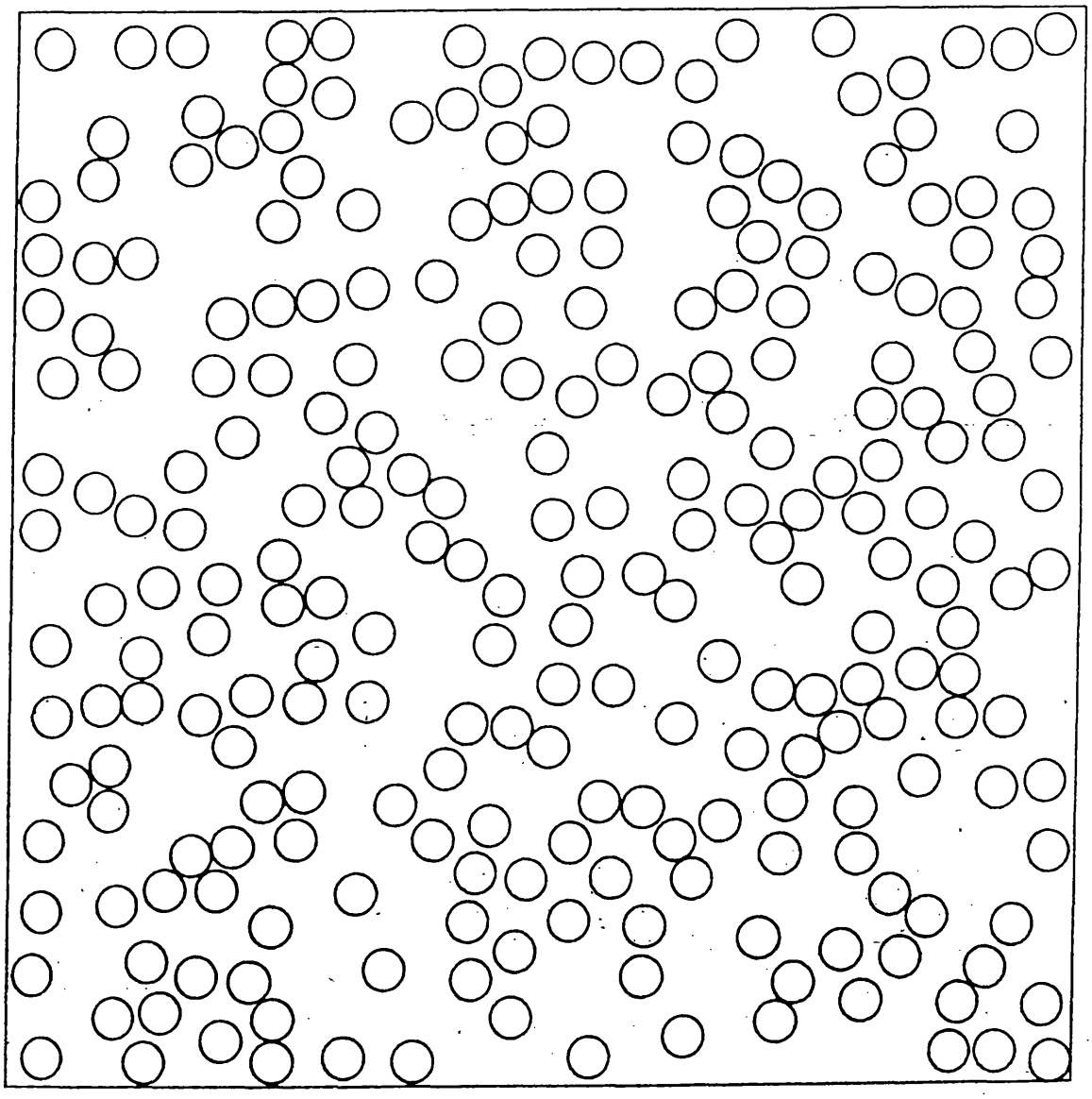
In the preceding sections, various coupling methods have been used to obtain convergence results for some spatial birth and death processes. In particular, ε-coupling was used to prove the convergence of the hard core fixed-number process for $N < 2(1 + R/2)(1 + \sqrt{3}R)/(3\sqrt{3}R^2)$ This limit on the packing density of the configuration is about one third of the theoretical maximum of 0.906; a configuration with N slightly below the limit is shown in Figure 1. Condition (5.11) was imposed in order to ensure that the statespace of the Markov chain formed a single class; while it is sufficient, it is not necessary, and the bound on the packing density could be raised.

However, for certain values of R and certain high packing densities, the statespace is not a single communicating class. Some examples are:

i) For $R = {}^{1}/_{n}$, $n \in \mathbb{N}$, and $N = (\frac{1}{R} + 1)^2$, it is possible to have a configuration with the centres of the discs forming a square lattice of side ${}^{1}/_{R}$, with lattice points at the corner of the square. The packing density here is ${}^{\pi}/_{4} = 0.7854$.

ii) For certain values of R, the discs can be arranged so that their centres form a grid of isosceles triangles of side R,R and R(1 + μ) (see Figure 2), for certain values of μ. Examples are $R = {}^{5}/_{12}$, $R = {}^{5}/_{36}$ and $R = {}^{1}/_{12}$, with μ = 0.6 in each case. In each case, the packing density is ${}^{\pi}/_{5.12} = 0.6136$. All the configurations above are such that each forms a single class - once a disc has been removed, it must be replaced in exactly the same position. Although each such configuration has zero probability of being entered, the above configurations do raise doubts about the convergence of the hard core birth and death process at very high packing densities. (Note that Ripley's finite statespace argument cannot be applied for packing densities admitting such configurations).

Figure 1 250 discs of diameter 0.04 with centres in the unit square. The square shown, drawn to include all the discs, has side of length 1.04.



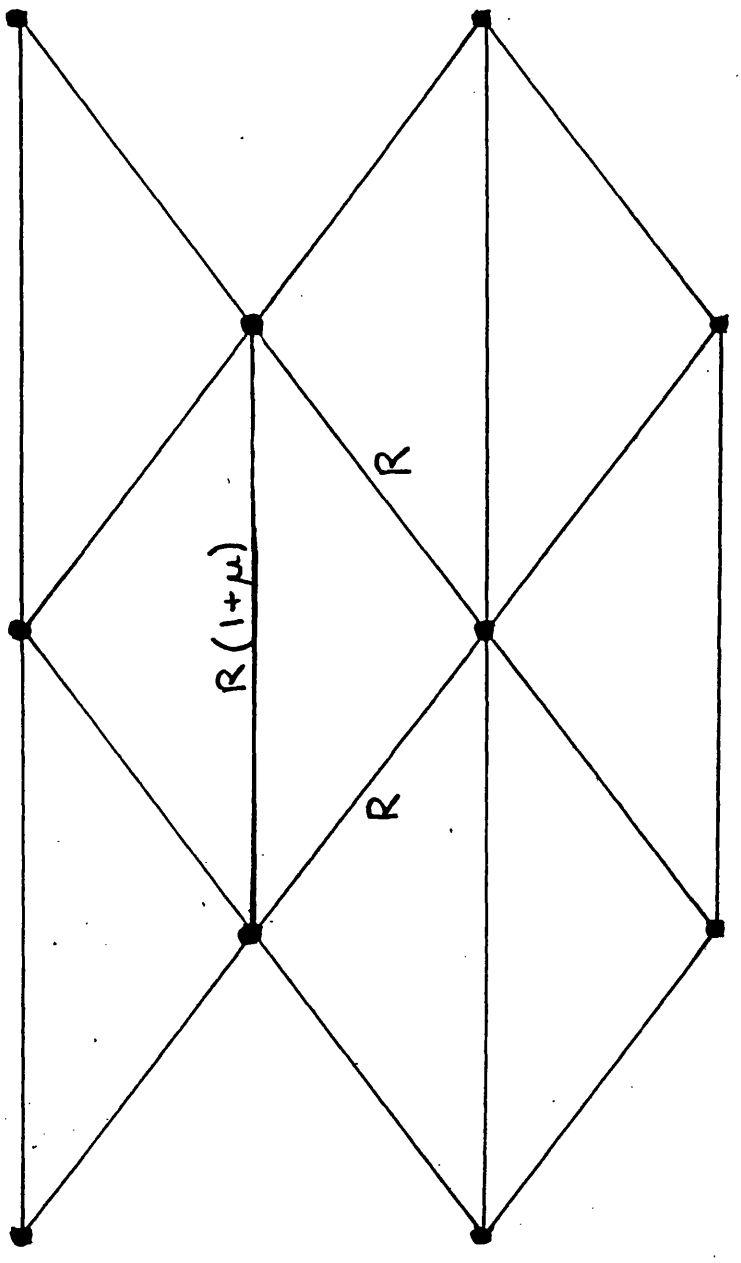


Figure 2 Arrangement of disc centres at the vertices of a grid of isosceles triangles, as described in section 9. In this example, $\mu = 0.6$.

In the absence of theoretical convergence results for high packing densities, an empirical study of the convergence properties at such densities would be helpful; this subject is treated in Chapter 6.

CHAPTER SIX

SIMULATING HARD CORE BIRTH AND DEATH PROCESSES

1. Introduction

It was noted in Chapter 5 that an empirical study of hard core birth and death processes at high packing densities could provide valuable insight into their convergence properties; such a study would, of course, be based on computer simulation of the processes. Unfortunately, the existing methods for simulating the processes, based on rejection sampling, were extremely slow at very high packing densities, making such an investigation impracticable. For this reason, a new algorithm, described in this chapter, has been written to simulate Kelly-Ripley hard core birth and death processes at high packing densities. This algorithm, which is based on the Dirichlet tessellation of the points, and makes use of the Green-Sibson TILE algorithm (Green and Sibson, 1978), is considerably faster than the old simulation methods at very high packing densities. This makes it possible to use the new algorithm in two contexts; firstly to investigate the convergence of the birth and death processes at high packing densities, and, secondly, to simulate Kelly-Ripley hard core models at high densities. Applications arise where hard core patterns with high packing densities are of interest; see, for example, Talbot and Willis (1980), and Alder and Wainwright (1962).

In section 2 of this chapter, the existing method of simulating the birth and death process is described, and the rationale behind the new algorithm is outlined. More details of the new method are

given in sections 3 and 4. In section 5, modifications to the algorithm are described which make it possible to simulate other hard core processes. The speed of the new algorithm is compared with that of the old methods in section 6. Finally, in section 7, some convergence properties of the birth and death processes are studied, using the new algorithm.

2. The old and new methods

The commonly-used method of simulating the Kelly-Ripley fixed number hard core birth and death process (described in Chapter 5, section 2b) is based on rejection sampling, and is described in Ripley (1979a). At the birth stage of each step, a new point is chosen uniformly in the sampling region. If it has none of the N-1 existing points within distance R of it, this point is accepted as the new point to be inserted; otherwise it is rejected, and another candidate for insertion is chosen uniformly in the sampling region, this process continuing until a suitable new point has been generated. The acceptability of a new point is determined by checking its distance from each of the N-1 existing points, and rejecting as soon as a distance less than R is found. For large N, the TILE algorithm of Green and Sibson (1978) can be used to find the shortest distance from the new point to an existing point; this should prove faster than checking the distances in turn. Whichever method is used, there will be many rejections for large R, since there will be very little available space for the insertion of a new point. As a result, this method is very slow for high packing densities, so that simulating at such densities is prohibitively expensive.

The new method uses very little rejection sampling, and is based on a different approach, so that it 'homes in' on the vacant space available for inserting a new point. At the birth stage of each step, the Dirichlet tessellation of the existing points is used to triangulate the sampling region in such a way that each triangle has one of the data points as a vertex. The construction of this triangulation is such that it is then a straightforward matter to compute how much vacant space there is in each triangle for the insertion of a new point. At high packing densities of discs, many of these triangles will contain no

vacant space, so that the list of triangles available for the insertion of a new point will not be very long. A triangle is then chosen at random from the list of available ones, with probabilities weighted according to how much vacant space there is in each triangle. This triangle having been selected, the new point to be born is inserted uniformly in the permitted part of the triangle, completing the birth stage of the step. The death part of each step is, of course, straightforward, and is done in the same way as the old method.

This new algorithm has been implemented in FORTRAN. The nature of the algorithm is such that it has a number of properties which make it efficient at high packing densities. These are given in the next section, where the new algorithm is described in detail.

3. Detailed Description of the Algorithm

The basic ideas of the algorithm were given in the previous section; here, in addition to describing it in more detail, we discuss some of its advantages over the rejection sampling method. We first give more details of the steps described in section 2.

i) Triangulating the square

This is done by connecting each point of the pattern to the vertices of its tile in the Dirichlet tessellation, as shown in Figure 1. Each triangle can be referenced by two pointers - the point of the pattern which is a vertex, and the neighbour which it is opposite (Figure 2). Note that if I and J are two points which are neighbours, then triangle(I,J)is a mirror image of triangle (J,I).

ii) Computing the amount of vacant space

The tile corresponding to a given point of the pattern consists of that part of the window closer to that point than to any other point of the pattern. It follows that the unavailable region of a triangle with point I as a vertex consists of the intersection of the disc of radius R, centred at point I, with that triangle (Figure 3). The calculation of the area of such an intersection is straightforward, and has already been described in Chapter 1 ; from this area we can, of course, obtain the area of the vacant region of the triangle.

iii) Keeping the list of available triangles

Information regarding the available triangles is kept in three arrays. The first of these records the point of the pattern which each available triangle has as a vertex; the second records the neighbouring point (or constraint) to which the triangle corresponds, and the third

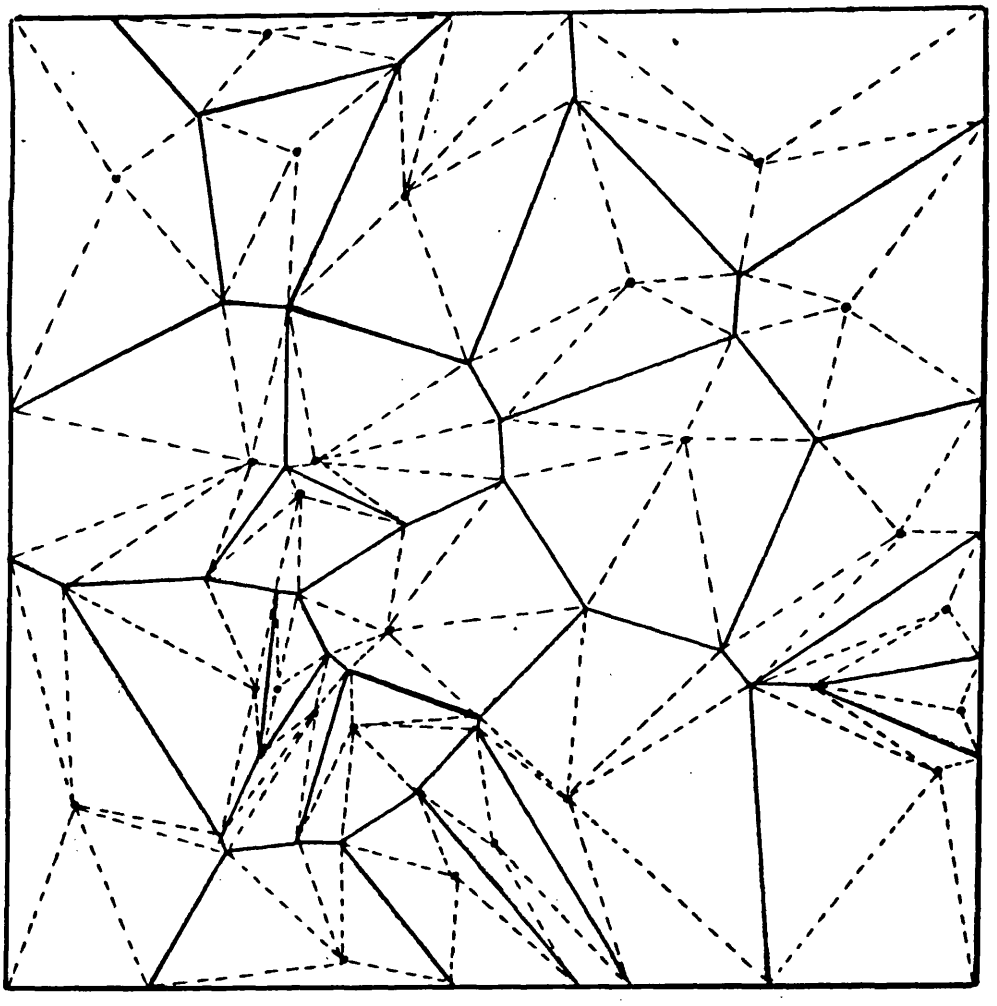


Figure 1 . Triangulation of the square using the Dirichlet tessellation. The unbroken lines are tile boundaries.

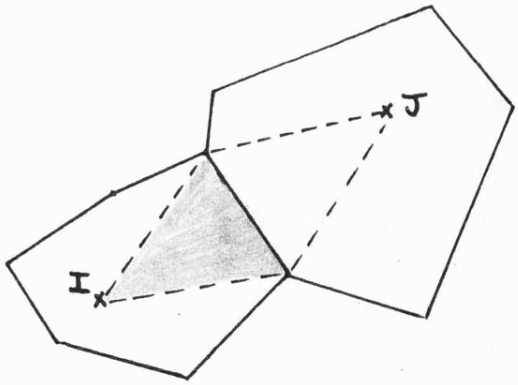


Figure 2 Referencing a triangle using the points of the pattern. The shaded triangle is triangle (I,J).

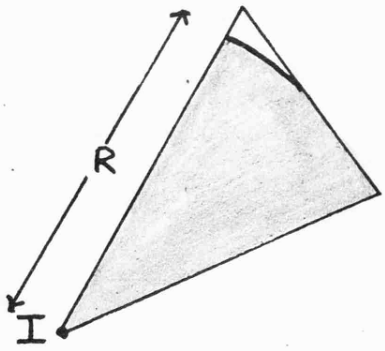


Figure 3 The unavailable region of the triangle shown is the shaded region.

records the area of the vacant region of the triangle. Triangles containing no vacant space need not be recorded. For neighbouring points I and J, the amount of vacant space in triangle (I,J) is the same as that in triangle (J,I), so that only one of these triangles need be stored; assuming that the points have been numbered, we store the triangle with the lower numbered point as a vertex.

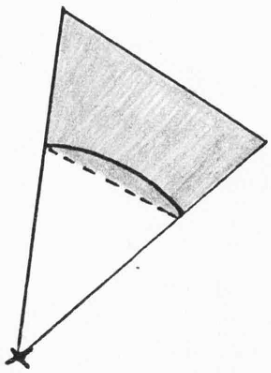
iv) Choosing an available triangle

To do this, we select at random from the available triangles, so that, if we have M available triangles with vacant areas $a_1, \ldots, a_n$, the probability of selecting the j^{th} triangle is $a_j/(a_1 + \ldots + a_m)$. This is simply sampling from a finite discrete distribution; we use a variant on the alias table method of Walker (1977). Further details of this step are given later in the section.

v) Inserting a new point uniformly in the vacant region of the chosen triangle

This is done by rejection sampling. A polygonal region P is constructed to contain the vacant region of the triangle (Figure 4); we place the new point uniformly inside P and use rejection sampling. The probability of rejection at this stage of the step is usually low.

Since the Dirichlet tessellation is a local construction, very little of the triangle structure is changed at each step of the birth and death process. Thus, after one step of the process, most of the triangles are unchanged, so that, rather than recompute the list of available triangles, we need only update it by adding new available triangles, deleting those which are no longer vacant and amending those whose vacant areas have changed. The number of changes to be made at each step is comparatively small, and does not increase with the number of points in the pattern. The list of available triangles is referenced



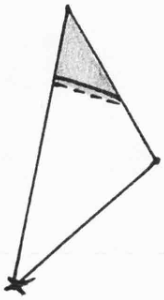


Figure 4 Examples of construction of a polygonal region P to contain the vacant region of the triangle. In each case, P is the shaded region.

from a heap similar to that used in the TILE algorithm (Green and Sibson, 1978), and the changes at each step are made in a way similar to that used in that algorithm.

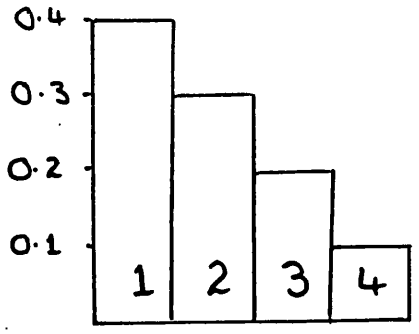
By using a modified alias table method for selecting the available triangle in which to insert the new point, we can make the algorithm more efficient by updating rather than recomputing the alias table after each step. Walker's alias method (Walker, 1977) works by expressing an n-point distribution as an equiprobable mixture of n two-point distributions, each of which is concentrated on one point and its "alias". This is illustrated diagrammatically in Figure 5. To select a point from the n-point distribution we first select one of the n two-point distributions; one comparison then suffices to choose between the corresponding point and its alias.

The modified alias table method

At the first step of the birth and death process, an alias table corresponding to the available triangles is set up, the probability corresponding to each available triangle being the proportion of the total available area lying in that triangle. The method for setting up the table is described in Kronmal and Peterson (1979). At subsequent steps, the alias table is modified as the available triangles change. These modifications introduce the possibility of rejection in choosing a triangle, increasing the time taken to select a triangle; when the probability of such a rejection becomes high, it is necessary to recompute the alias table at some stage and then proceed with modifications in subsequent steps. The method of modification of the alias table, illustrated in Figures 6 and 7 is as follows. Two types of modification are carried out at each stage:

Figure 5 Diagrammatic representation of the alias method. The weight in the bar chart representing an n-point distribution is redistributed to form n bars, each of height $1/n$. To select a point, we first select a bar, i; then with probability f_i, the point chosen is i, otherwise, its alias, A_i is selected.

This is illustrated below with n = 4 and a distribution with probabilities $p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.2$ and $p_1 = 0.1$. The alias table constructed has $A_1 = 2$, $A_2 = 2$, $A_3 = 1$, $A_4 = 1$, $f_1 = 0.8$, $f_2 = 1.0$, $f_3 = 0.8$, $f_4 = 0.4$.



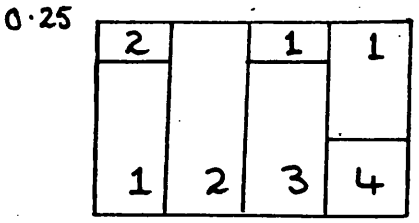
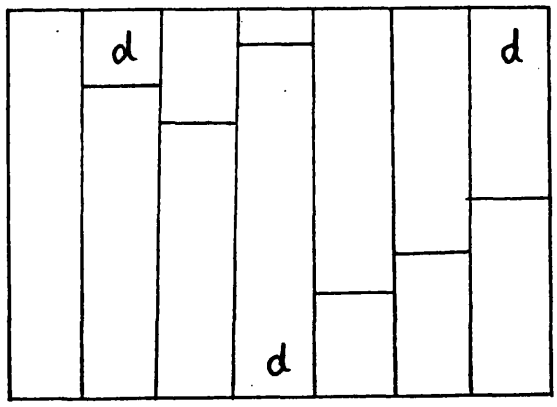
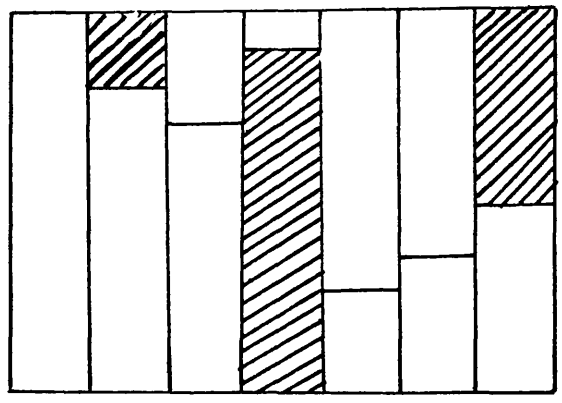


Figure 6 Removal of a triangle from the alias table.

Original table:



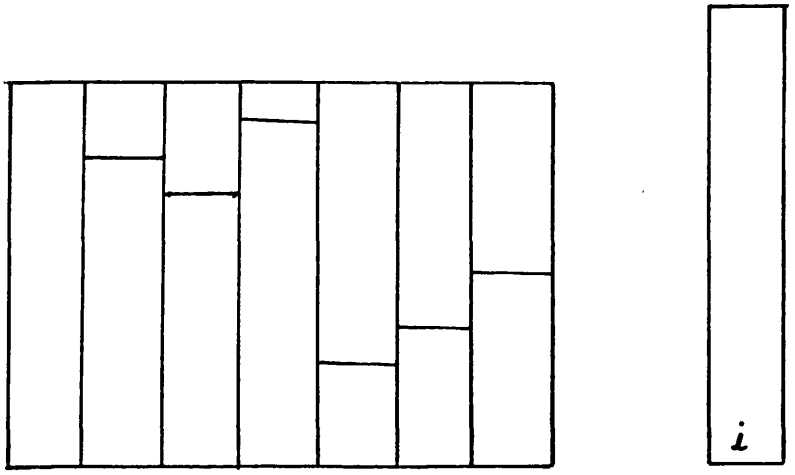
Removal of triangle d leaves:



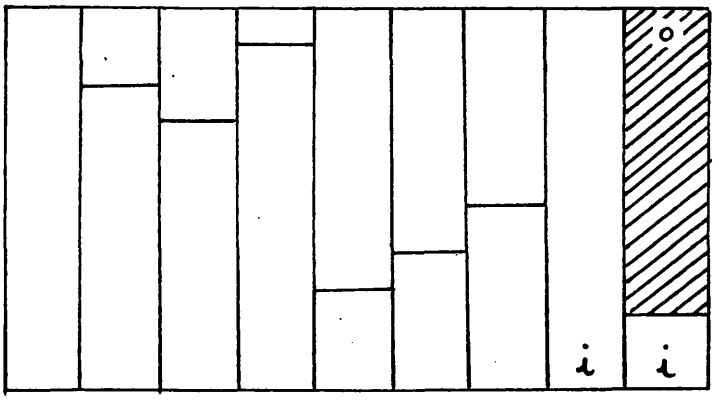
Rejection occurs if a shaded region is chosen.

Figure 7 Addition of a triangle to the alias table

Original table and triangle to be added:



Addition of the new triangle yields



i) Removal of a previously available triangle.

This is done by effectively deleting the parts corresponding to the triangle from the alias table (see Figure 6); thus, if the triangle is chosen, a rejection occurs and the process of choosing a triangle is restarted.

ii) Addition of a new triangle

For this modification, extra 'bars' are added to the table to allow for the new weight; part of the final bar will be blank, and is indexed by a dummy triangle 0, a rejection occurring if triangle 0 is chosen. This corresponds to aliasing this bar with triangle 0; see Figure 7.

Triangles in which the available area has changed during the step are first deleted and then added to the table with their new areas; thus only the two operations above are necessary. At each step, a number of removals and additions will be made.

Suppose that, before an operation, there are N bars in the table, the total area accounted for by the table is TA, and the proportion of the table corresponding to rejection is PREJ.

Removal of a triangle with previously available area A brings about the following changes:

N is unchanged

Total area becomes TA - A

Rejection proportion becomes $\text{PREJ} + \frac{A \times (1 - \text{PREJ})}{TA}$

Addition of a triangle with available area A changes the table in the following way:

$$N \text{ becomes } N + \left[\frac{A \times N(1 - \text{PREJ})}{TA} \right] + 1 \quad .$$

Total area becomes TA + A

Rejection proportion becomes

$$1 - \frac{N \times (1 - PREJ)}{N + 1 + \left[\frac{N \times A \times (1 - PREJ)}{TA}\right]} \left(1 + \frac{A}{TA}\right)$$

Selection of a triangle using the modified alias method requires three comparisons: two as before to select a possible triangle, and one further comparison to test for rejection. If rejection occurs, the process is repeated. After many modifications, the rejection probability will become high, and it will then be desirable to recompute rather than modify the table at the next step; i.e. set up a fresh alias table from the current triangle distribution. The question of how often the table should be recomputed is discussed in section 4.

The logic of the new algorithm is illustrated by the flow diagram in Figure 8.

The new algorithm uses rejection sampling only in the alias table and in inserting the new point in the permitted part of the chosen triangle. The probability of rejection can be controlled in the first case (by recomputing the alias table), and in the second case is comparatively low. These facts make the new algorithm considerably faster than Ripley's (1979a) at high packing densities. As the packing density increases, the available area becomes smaller, making the list of available triangles and the alias table smaller, so that the speed of the new algorithm should increase. These trends are illustrated by the timing runs in section 6.

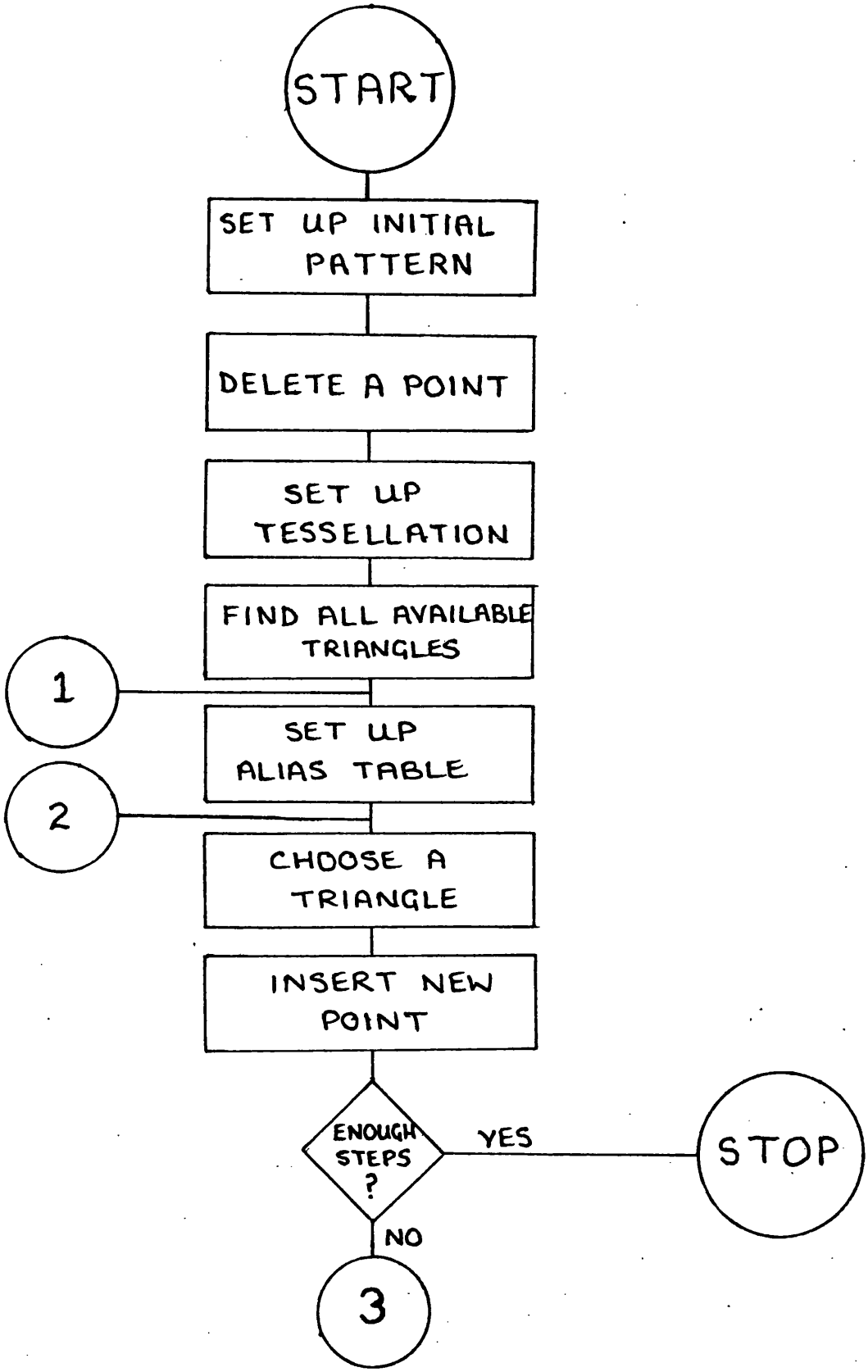


Figure 8 Flow diagram for the new simulation algorithm.

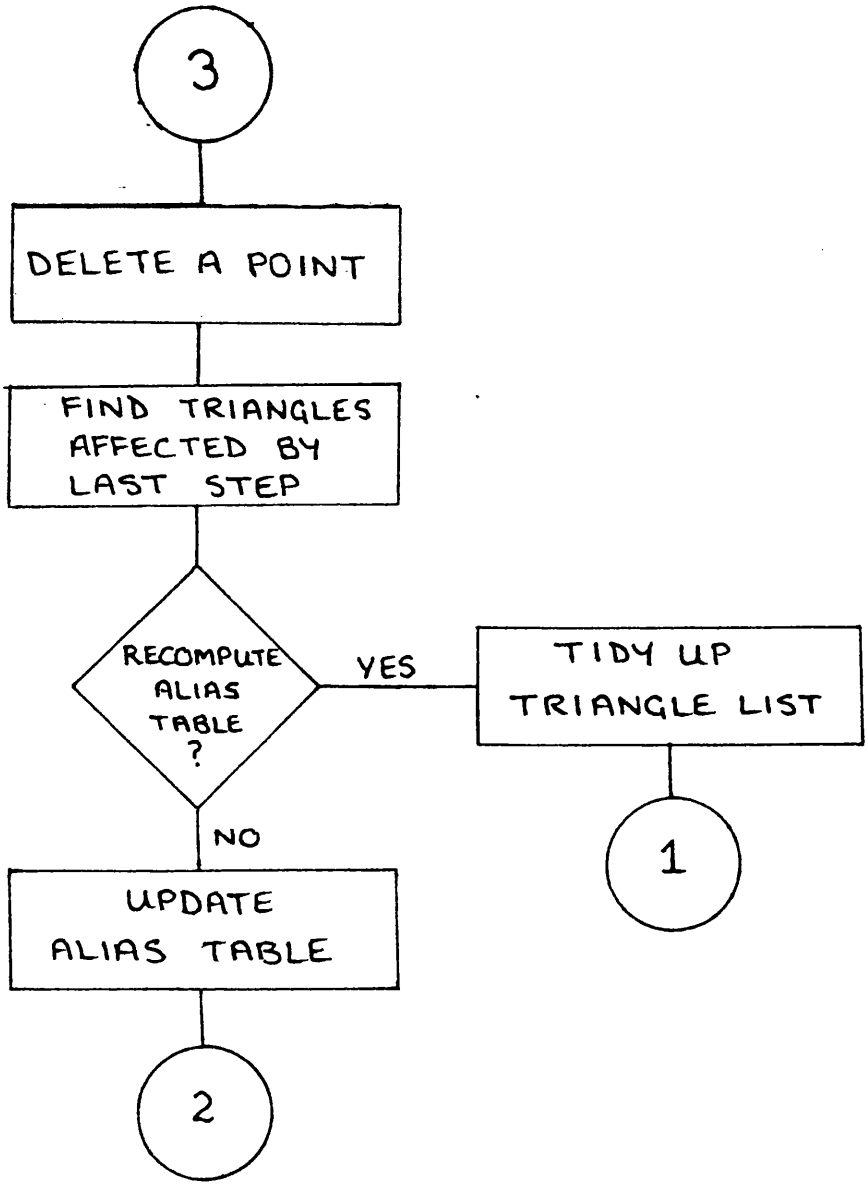


Figure 8 (Continued)

4. Deciding when to Recompute the Alias Table

As was noted in section 3, it is necessary to recompute, rather than modify, the alias table when the rejection probability becomes too high; we therefore need a criterion for deciding when the table should be recomputed.

We can attempt to formulate the problem of deciding when to recompute the alias table in Dynamic Programming terms. When this is done, the problem is seen to be similar to a 'machine replacement problem' (see Ross, 1970), where, for our purposes, replacing the machine corresponds to recomputing the table.

After the n^{th} step of the birth and death simulation, let N_n be the number of bars in the table, let TA_n the total area accounted for by the table, and let $PREJ_n$ be the proportion of the table corresponding to rejection. We consider the stochastic process $T_n = (N_n, TA_n, PREJ_n)$. At any step of the simulation, we have to decide whether to modify or recompute the alias table after the next step. The result of either course of action will, of course, depend on the outcome of the next step.

If the current state is (N,T,p), denote by (N',T',p') the state attained by modifying the table, and by (N",T",O) the state attained by recomputing the table. To formulate our problem as a Markov decision process, we make the following simplifying approximations:-

i) $(N_n, TA_n, PREJ_n)$ is a Markov chain.

ii) (N", T") is independent of (N, T, p).

These assumptions would appear to be reasonable approximations to the behaviour of the process, making a little theoretical progress possible; the complex nature of the inter-relation between the birth

and death process and the alias table makes it difficult to progress without making numerous approximations. Our objective is deciding whether or not to recompute the alias table is to minimise the time taken to select a triangle. Thus, when we consider the problem from the point of view of Dynamic Programming, the 'costs' will be expressed in terms of the time taken to perform various operations. Let B be the time taken to select a bar in the table (this will not depend on the size or state of the table), let R be the expected time taken to recompute the table, and S the expected time taken to modify the alias table.

Each time the process is in state (N, T, p), an expected cost C(N, T, p) is incurred. For $p > 0$, this expected cost is $\frac{B}{1 - p} + S$; for $p = 0$ it is $B + R$. (Note that, here, we are including the cost of modifying or recomputing the table in the cost incurred at the following step). The problem of deciding when to recompute the table is similar to the Markovian replacement problem considered by Ross (1969), the main difference being that the statespace of our Markov process is multidimensional, rather than being a subset of $\mathbb{R}$. Ross considers the following problem: a unit is observed at the beginning of discrete time periods and classified into one of a number of states labelled by some Borel subset of $\mathbb{R}$. After observing the unit, the observer must choose one of two possible actions: to leave the unit in service or to remove the unit from service at the end of the period. If the latter action is taken, a new unit is purchased, and the system restarts in state '0'. Ross considers both the discounted cost and the expected average cost criteria, and, for each, gives conditions under which there is an optimal policy which is of the form "replace if the state x is greater than a threshold value x_o". The main condition for this to hold is that the conditional probability of a transition into

any block of states (y,∞) given that the unit is left in service is a non-decreasing function of the present state x (see Ross (1969) for more details).

For our problem, of course, replacing a unit corresponds to recomputing the alias table. The expected costs C, which depend only on the rejection probability p, are, in theory at least, unbounded, so that one of Ross's conditions is immediately violated.

The rejection probability p does, however, correspond in some ways to the 'state' x in Ross's formulation.

While Ross's results cannot be applied directly to this problem, they do provide some backing for the use of the intuitively appealing decision rule 'recompute if $p > p_o$', where p_o is some threshold value. This is the form of rule which has been used in the simulation algorithm described in this chapter. Experimenting with the value of p_o led to the choice of 0.3 as the threshold value. The optimum value of p_o would, presumably, depend on the number of points in the birth and death process and on the diameter of the discs; since, however, empirical investigation suggested that varying p_o between 0.2 and 0.4 made little difference to the overall time taken (see Table 1 for a particular example), the threshold value was kept fixed at 0.3 for all simulations. Recomputing or amending the alias table are comparatively inexpensive operations, and the above decision rule has proved adequate in practice.

TABLE 1

Average times taken on an ICL 2980 computer to simulate 1000 steps of the fixed number process for various values of the threshold parameter p_o. Ten simulation runs were performed for each value of p_o, using 200-point configurations with R = 0.06; the initial configuration positioned the points at the vertices of a square grid of side 0.06001.

p_o	Time (in seconds)
0.2	40.32
0.25	39.57
0.3	39.16
0.35	40.39
0.4	38.45

5. Other Applications

The algorithm described in sections 2 and 3 can easily be adapted to simulate other hard core patterns.

i) The General Hard Core Birth and Death Process

This is described in section 2 of Chapter 5. The jump chain of the process can be simulated by adapting the new algorithm; rather than births and deaths alternating, the probability at any stage that the next event is a death is $\frac{m}{m + k\beta}$, where m is the number of points in the pattern, k is a constant, and β is the total area available for inserting a new disc. This last quantity can be found by summing the area available in each triangle; a single comparison is then used to decide whether the next step should be a birth or a death. The quantity β need not be evaluated from first principles at each step, but can be updated from step to step.

ii) The SSI Process

This is described in section 1 of Chapter 7; here, there are no deaths, but discs are sequentially inserted until the required number of discs has been inserted. The new algorithm is easily adapted to simulate the SSI process; details are given in section 2 of Chapter 7.

6. Timing Runs

In this section, we compare the times taken by the old and new algorithms to simulate the birth and death process at various packing densities. Three methods are compared:

i) The rejection sampling method outlined in section 2 and described by Ripley (1979a). (Method R1)

ii) The rejection sampling method which uses the TILE algorithm of Green and Sibson (1978) to find the shortest distance from the new point to an existing point. (Method R2)

iii) The new method described in section 4. (Method T).

Table 2 shows the time taken per step of the process for patterns of 200 points in the unit square, at various packing densities. The timing runs were carried out using a regular starting pattern, with the points at the vertices of a square grid. Most of the runs lasted for 3000 steps; the time taken by the rejection sampling methods at high packing densities meant that shorter runs had to be used in some cases. The times are those taken using an ICL 2980 computer. The information in Table 2 is represented graphically in Figure 9.

The time taken by the new algorithm varies little; for lower packing densities it is much slower than both the old methods, but is faster for packing densities greater than 0.45, and, for the very high packing densities, is the only practicable algorithm to use. Algorithm R2 is always slower than R1; for larger numbers of points, we might expect it to be better in some cases.

Table 3 and Figure 10 show the results of similar timing runs for patterns of 400 points in the unit square. Again, most of the runs were for 3000 steps, with a regular starting pattern. The results are

TABLE 2

Time taken per step to simulate patterns of 200 points at various packing densities. All timing runs, apart from those marked *, lasted for 3000 steps.

R	Packing Density	Time per step (seconds)		
		Algorithm R1	Algorithm R2	Algorithm T
0.025	0.093	0.0041	0.0183	0.0527
0.030	0.133	0.0044	0.0183	0.0526
0.035	0.180	0.0048	0.0186	0.0521
0.040	0.232	0.0055	0.0191	0.0519
0.045	0.291	0.0073	0.0201	0.0504
0.050	0.356	0.0115	0.0225	0.0460
0.055	0.427	0.0241	0.0326	0.0412
0.060	0.503	0.0897	0.0939	0.0340
0.065	0.585	0.4632	0.7895	0.0316
0.070	0.672	1.9640*	2.0145*	0.0318

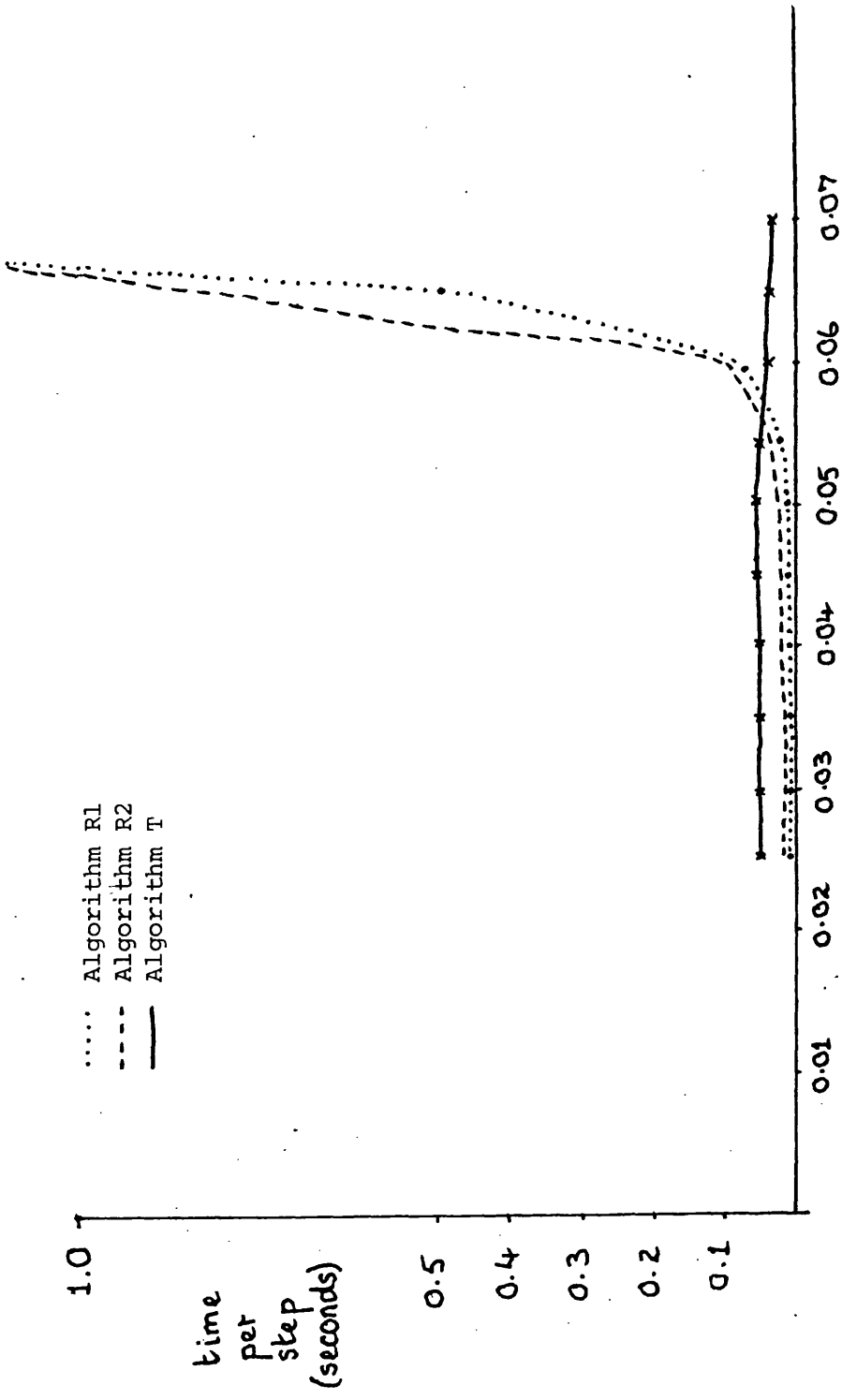


Figure 9 Time taken per step to simulate patterns of 200 points at various packing densities.

TABLE 3

Time taken per step to simulate patterns of 400 points at various packing densities. All timing runs, apart from those marked *, lasted for 3000 steps.

R	Packing Density	Time per step (seconds)		
		Algorithm R1	Algorithm R2	Algorithm T
0.010	0.031	0.0055	0.0176	0.0768
0.015	0.069	0.0057	0.0181	0.0768
0.020	0.121	0.0062	0.0184	0.0768
0.025	0.187	0.0073	0.0184	0.0732
0.030	0.267	0.0099	0.0195	0.0695
0.035	0.359	0.0181	0.0234	0.0693
0.040	0.465	0.0771	0.0480	0.0431
0.045	0.583	1.7647*	0.5916	0.0314
0.050	0.712	81.081*	56.604*	0.0322

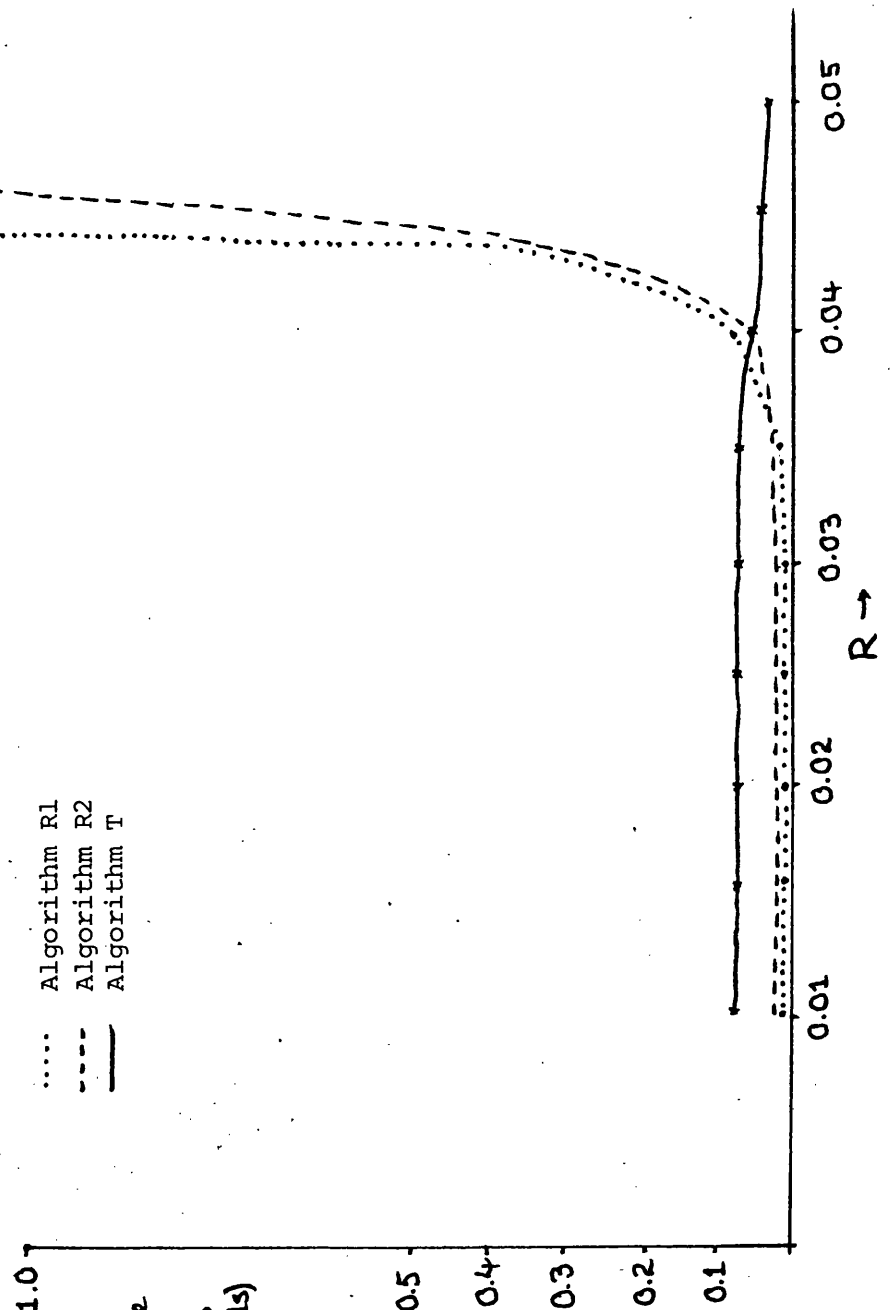


Figure 10 Time taken per step to simulate patterns of 400 points at various packing densities.

similar to those for 200 points, with the new algorithm again being faster than the old ones for packing densities above 0.45, and considerably faster for high packing densities. Algorithm R2 is again slower than R1 for the lower packing densities, but is faster for packing densities above 0.4

The results of this section suggest that it is best to use the new algorithm for packing densities of 0.45 or over; for packing densities above 0.5, use of the other two algorithms would appear to be impracticable.

7. Convergence of the fixed number process at high densities

We investigate the convergence of the fixed number process by studying the behaviour of some statistics of the process.

Let $\{X_n\}$ be a (simulated) fixed number birth and death process in the unit square and let S be a statistic on the space of point patterns. For fixed $\ell \in \mathbb{N}$, write

$$S_m = S(X_{\ell m}) \qquad m \in \mathbb{N}$$

We can regard $\{S_m\}$ as a time series, which will be stationary when $\{X_{\ell m}\}$ is in equilibrium. We study the convergence of the birth and death process by looking for stationarity in the time series $\{S_m\}$. To do this, we need to have a method of deciding when $\{S_m\}$ has 'settled down' to stationarity. We do this using cumulative sum techniques (see, for example, Woodward and Gold-Smith (1964) or Wetherill (1969)).

Suppose we simulate the process, starting from a regular grid of points, for $M\ell$ steps, where M is large. Let

$$T_m = S_{M-m+1} \qquad (1 \leqslant m \leqslant M).$$

Figure 11 shows a plot of S_m for a realisation of a process with 200 points and R = 0.055, for M = 500, ℓ = 20, and using as our test statistic the empty space statistic of Chapter 1 (without edge correction) for discs of radius 3R/2. To estimate the point where S_m becomes stationary, we estimate the point where T_m ceases to be stationary. Figure 12 shows a cusum plot of the series T_m derived from that of Figure 11; the 'target value' t was the mean of the 500 observations in the series. Formally, we define,

$$U_m = \sum_{i=1}^{m} (T_i - t)$$

$$t = \frac{1}{500} \sum_{j=1}^{500} T_j ,$$

and plot U_m against m.

Changes of mean in the series T_m will be revealed as changes of slope on a cusum chart. Bearing this in mind, visual inspection of Figure 12 enables us to estimate as m = 145 the point at which T_m ceases to be stationary. This gives us an estimate of r = 7100 for the point at which the statistic $S(X_r)$ "settles down" to stationarity, indicating that the birth and death process itself is close to equilibrium by this stage. An automatic method for estimating the point at which the series $\{T_m\}$ ceases to be stationary is provided by the use of the 'post-mortem' technique described in section 5.3 of Woodward and Goldsmith (1969), and in 12.10.2 of Chatfield (1975). Firstly, the residual variance of the series $\{T_i\}$ is estimated by

$$s^2 = \frac{1}{2(n-1)} \sum_{i=1}^{M-1} (T_{i+1} - T_i)^2 .$$

A Student's t-test is then used to locate turning points in the cusum chart, in the following way. We move along the series $\{U_m\}$, connecting each observation with the last turning point (or, initially, with the first observation), and find the maximum absolute distance between this chord and the intervening cusum readings. The position of this largest difference is used to divide the intervening original observations (the T_m's) into two groups, the point of largest difference being included in the first group. A t-test is then carried out between the mean

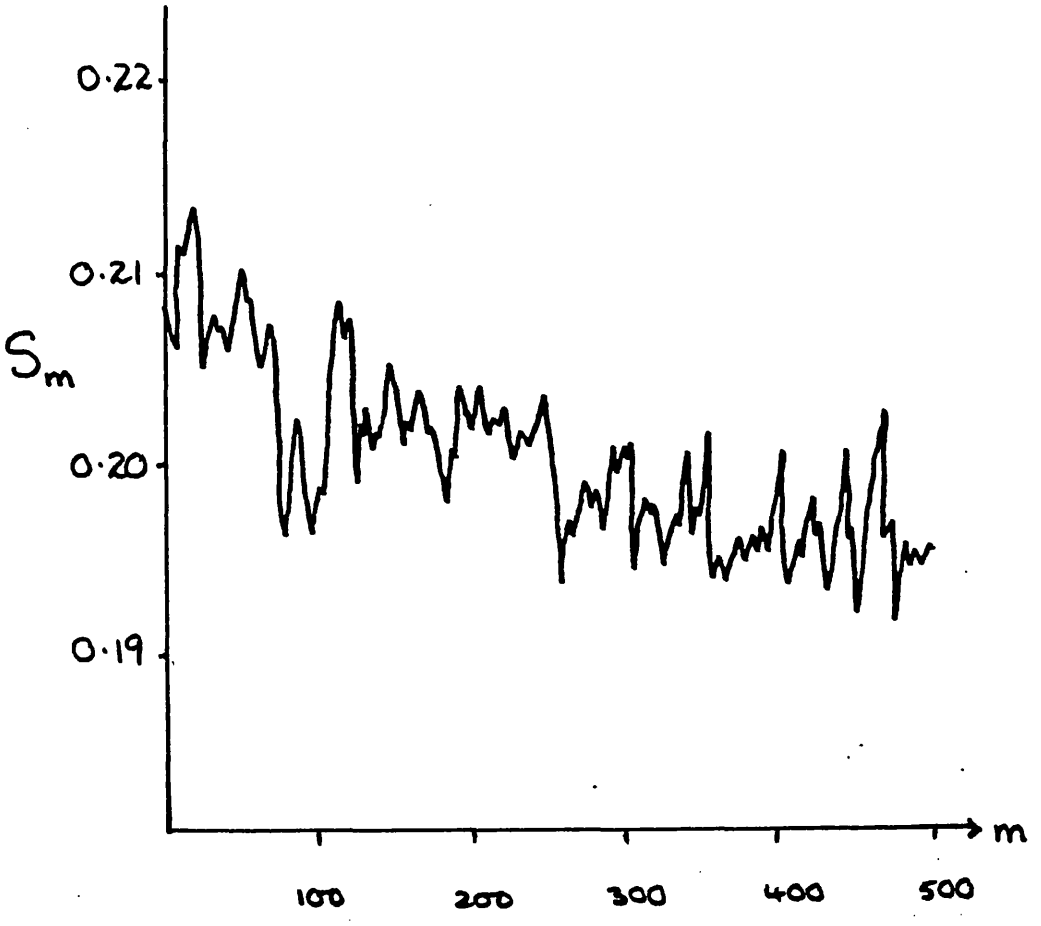


Figure 11 A plot of S_m for a realisation of a hard core process with 200 points and R = 0.055, for M = 500, ℓ = 20 (see text). The statistic used is the empty space statistic, without edge correction, for discs of radius 3R/2.

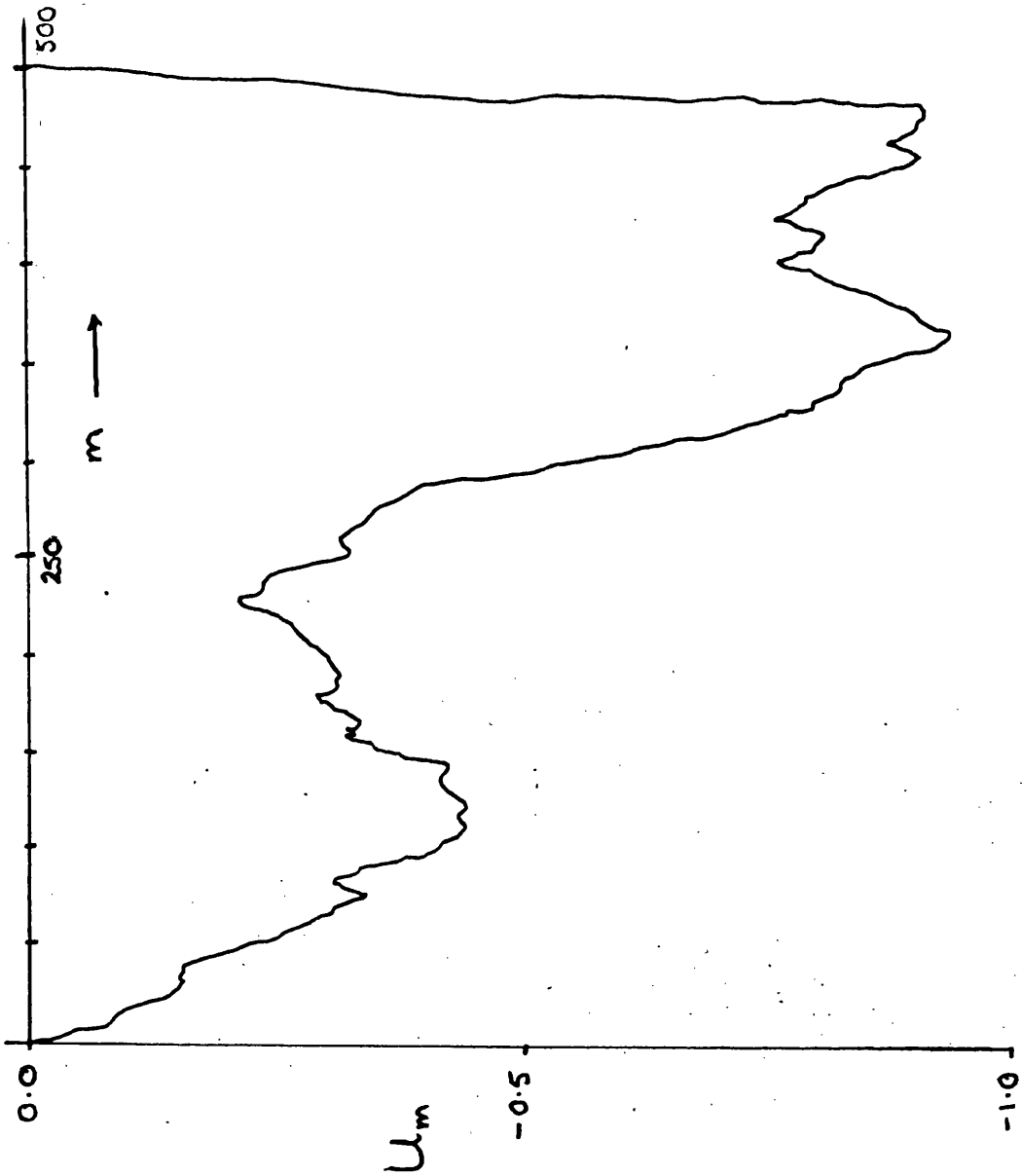


Figure 12 Cusum plot derived from the series of Figure 11.

values of the two groups using the estimated variance given above. When the value of t is not significant , we move to the next cumulative sum and repeat the process. When the value of t is significant we assume that we have found a turning point, at the point of largest difference.

Although our observations are not independent and are not known to be normally distributed, the t-test described above does serve to identify changes of slope in the cusum chart. We look for the first large change of slope in the $\{U_m\}$ chart, and take the position of this change as our estimate of the point at which $\{T_m\}$ ceases to be stationary. To avoid locating spurious changes, we apply a stringent condition, and require the absolute value of the t-statistic to be above a threshold value corresponding to a significance level of 0.05% before we assume that we have found a turning point.

Application of this procedure to the series of Figure 12 gave an estimate of m = 147 as the point at which T_m ceased to be stationary; this result is close to that obtained by visual inspection of the cusum plot. This procedure was carried out for a number of simulation runs. Patterns of 200 points were used each time, with 20 runs being carried out for each of four values of the disc diameter R : R = 0.05, 0.055, 0.06 and 0.065. An initial configuration with the disc centres positioned at the vertices of a square grid of side R + 0.0001 was used for these simulation runs. Two statistics were used during each run: the empty space statistic (without edge correction) for circular test sets of radius 3R/2, and the length of the boundary of the union of the discs of radius 3R/2 centred at the points of the pattern; as was noted in section 5 of Chapter 1, this latter statistic is related to the derivative of the empty space area. For each run and each test statistic, the point at which T_m ceased to be stationary was estimated. Precautionary tests were also

carried out over the last 1000 steps of the birth and death process to check that the series $\{S_m\}$ was stationary; no evidence of non-stationarity was obtained. The results of the simulation runs are summarised in Table 4.

As one would expect, the estimated time taken for the birth and death process to converge to equilibrium increases with the disc diameter R. The process appeared to have converged to equilibrium by the end of each run, but the convergence was slow. The results of Table 4 suggest that, when simulating Kelly-Ripley hardcore models at high packing densities such as those considered here, the process should be allowed to run for many thousands of steps before the pattern is used as an approximation to a realisation of the required process.

TABLE 4

Mean estimated time to convergence for hard core birth and death processes of 200 steps, at various packing densities, using two statistics:

S1 - Empty space statistic, without edge correction, for discs of radius 3R/2.

S2 - Length of the boundary of the union of the discs of radius 3R/2 centred at the points of the pattern.

(Figures in brackets are the standard errors of the mean estimates).

	Mean estimated time to convergence (steps)	
R	S1	S2
0.05	5839 (393.6)	5609 (358.2)
0.055	7053 (321.7)	6441 (353.8)
0.06	7404 (255.0)	7428 (272.4)
0.065	7509 (197.8)	7622 (200.6)

CHAPTER SEVEN

SIMULATING THE SSI PROCESS, AND THE COMPLETE PACKING PROBLEM

1. Introduction

The simple sequential inhibition (SSI) process was introduced by Diggle, Besag and Gleaves (1976) as a model for regular point patterns in a finite region. Discs of a fixed diameter R are placed sequentially over a finite region of area A. At each stage, the next disc centre is chosen uniformly at random from those points which ensure that no two discs overlap. The procedure terminates when a prescribed number N of discs are in position, and the final pattern is then defined by the disc centres.

When the packing density $\frac{N\pi R^2}{4A}$ is large, a stage may be reached when fewer than N discs have been inserted but there is no space available for the insertion of another disc. The highest packing density which we can guarantee to attain without this situation arising is, in the limit, 0.22672 (corresponding to dense hexagonal packing by discs of diameter 2R).

For relatively high packing density, methods of simulating the SSI process based on rejection sampling will be slow at the later stages of simulation, when there is little space available for the insertion of another disc. The algorithm described in Chapter 6 can be modified to simulate the SSI process; since it is not based on rejection sampling, the algorithm has no difficulty in simulating patterns with such packing densities. The modification of the

algorithm is described in section 2.

In introducing the SSI process, Diggle, Besag and Gleaves (1976) noted that it may be interpreted as a two dimensional analogue of the "car-parking problem" (see Mannion, (1964)). No theoretical results are available concerning the distribution of the number of discs that can be inserted under the SSI model before complete packing is attained, i.e. before there is no more space available for the insertion of another disc. In section 3 of this chapter, the new algorithm of section 2 is used to estimate the expected value of the packing density at complete packing, and density estimates of the packing density at complete packing are obtained.

2. Simulating the SSI process

Modification of algorithm T of Chapter 6 to simulate the SSI process is straightforward; simply start with no points, and omit the death part of each step. A fixed number of steps can be carried out using rejection sampling; the list of available triangles is then set up, and subsequent steps carried out using the new algorithm. The results of the timing runs of section 6 of Chapter 6 can be used to suggest the following simple rule: use rejection sampling until the packing density exceeds 0.4; thereafter use the new algorithm until the desired number of discs has been inserted. Since the Dirichlet tessellation of the points must be computed at some stage, for use with the new algorithm, we compute it at the beginning, and carry out the rejection sampling using the TILE algorithm of Green and Sibson (1978) to find the shortest distance from the new point to an existing point.

Since the new algorithm keeps a record of the available space for the insertion of a disc, it can be used to continue inserting discs until there is no more space available - this happens when there are no available triangles. This is a further advantage over the rejection sampling method. This property of the new algorithm means that it can be used to investigate the complete packing problem for discs in two dimensions. This is done in the next section.

3. Random sequential packing of discs in two dimensions

In this section, we consider random sequential packing of non-overlapping discs in a rectangular region in two dimensions, and use the algorithm described above to tackle the random complete packing problem for discs.

Random sequential packing has been considered by many authors. Non-overlapping objects are sequentially inserted in a container of finite size, the centre of each object being uniformly distributed around occupiable space of the container, i.e. over those points which ensure that no two discs overlap. This process continues until there is no space available for the insertion of a disc. Most of the theoretical work has been concerned with one-dimensional problems (see, for example, Renyi (1958), Page (1959) and Mannion (1976)); in higher dimensions, Palasti (1960), Blaisdell and Solomon (1970), Akeda and Hori (1975 and 1976), and Jodrey and Tory (1980), among others have considered random packing of n-dimensional cubes. (See Tanemura (1979) for further references). Much attention has been concentrated on the limiting packing density; that is, the limit as the container size tends to infinity of the ratio of the total volume of the packed objects to the container volume.

Tanemura (1979) provided a computational technique for a random complete packing of two-dimensional discs, and used this to give the value 0.5473 for the mean of the limiting packing density of discs; this was an improvement on the result of 0.4756 obtained by Solomon (1967). Tanemura's simulation technique involves an approximation in that, at its later stages, the discs are not inserted uniformly over the occupiable space. Rather, Tanemura subdivides the available space into a number of occupiable regions. These regions are selected in order for the insertion of a disc centre, irrespective of their area.

The algorithm of section 2 does not have this disadvantage, in that it continues to insert discs uniformly until complete packing has been attained; it should therefore give a more accurate estimate of the limiting packing density.

Simulations were carried out using the algorithm for eight container sizes. Discs of unit diameter were used, and the container sizes were chosen to correspond to those used by Tanemura (1979); each rectangle entirely encloses a system of hexagonally close packed discs and is almost square. The results of the simulations are shown in Table 1; for comparison, Tanemura's results are also shown. We denote by ρ_{xy} the packing density obtained at complete packing in a rectangle of sides of length x and y. As Table 1 shows, the values of the mean packing density $\bar{\rho}_{xy}$ obtained using the algorithm are consistently higher than those of Tanemura. The simulations summarised in Table 1 were used to estimate the probability density function of ρ_{xy} for the eight values of (x,y) used. The method used was the kernel method, using a Gaussian kernel (see Silverman, 1982), the estimates being computed using the Fast Fourier Transform algorithm of Silverman (1982). The choice of the window width in each case was determined using the 'test graph' method of Silverman (1978a). This involves drawing graphs (called test graphs) of the second derivative of the density estimate for various window widths, and choosing the window width corresponding to a test graph whose fluctuations are quite marked but do not obscure the systematic variation completely. The principle is illustrated in Figure 1, which shows test graphs for estimating the density function of $\rho_{11.0 \times 11.392}$; the test graphs for window widths 1.4 and 1.0 are respectively too smooth and too rough, while a window width of 1.2 appears to be acceptable.

TABLE 1

Mean packing density, $\overline{\rho}_{xy}$, obtained at complete packing of discs of unit diameter in a rectangle of sides of length x and y, using the new algorithm. The results of Tanemura (1979) are also given, for comparison.

	Tanemura			New Algorithm		
x×y	Number of trials	$\overline{\rho}_{xy}$	s.d. of $\overline{\rho}_{xy}$	Number of trials	$\overline{\rho}_{xy}$	s.d. of $\overline{\rho}_{xy}$
11.0×11.392	100	.51580	.01439	100	.52184	.01583
16.0×16.589	100	.52504	.01051	100	.53058	.01027
21.0×21.785	50	.52957	.00761	100	.54087	.00751
31.0×32.177	50	.53479	.00543	100	.54689	.00512
41.0×42.569	50	.53901	.00385	100	.54911	.00382
51.0×51.230	50	.54072	.00278	100	.55085	.00308
61.0×61.622	40	.54134	.00253	100	.55201	.00259
81.0×82.406	20	.54380	.00176	50	.55408	.00190

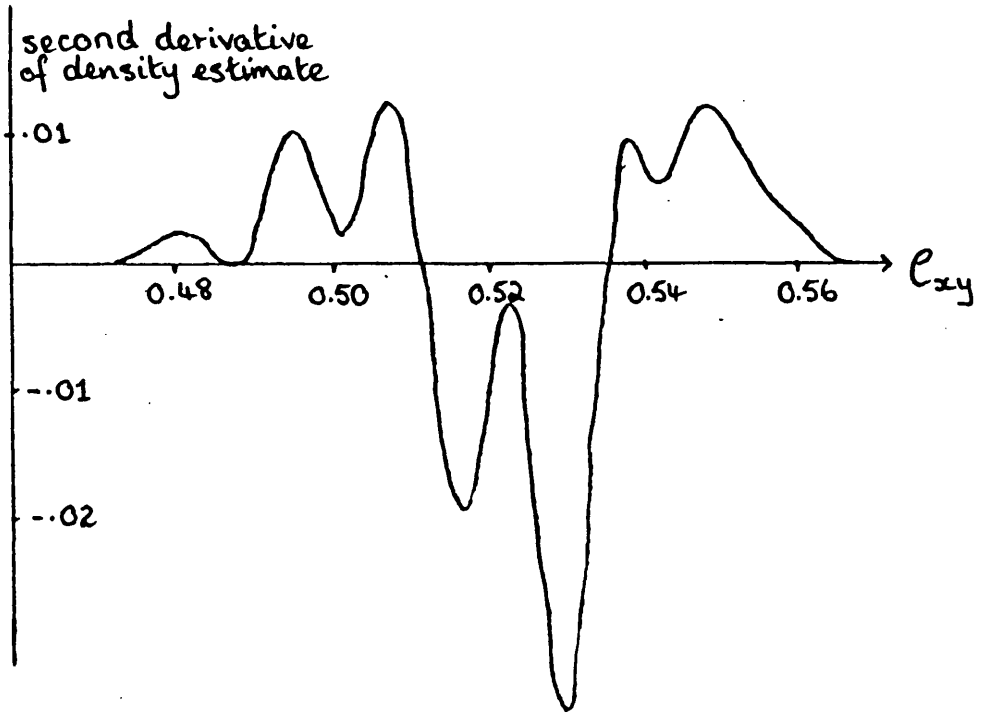
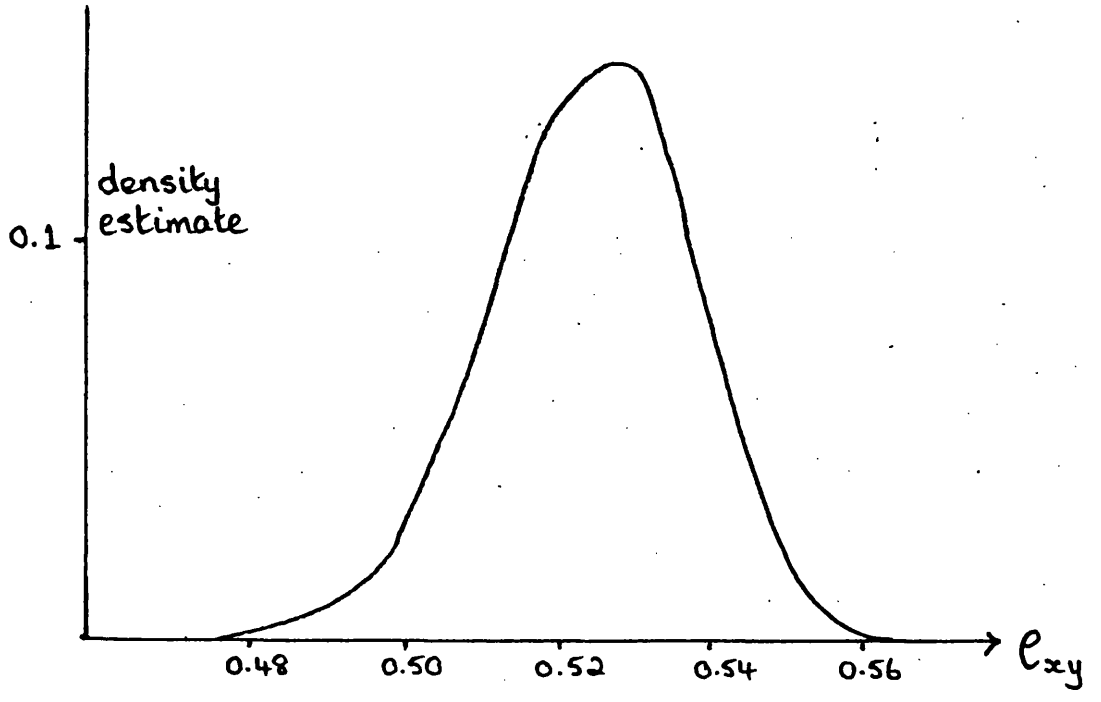


Figure 1a Test graph and density estimate at window width 1.0 for the density function of $\rho_{11.0 \times 11.392}$



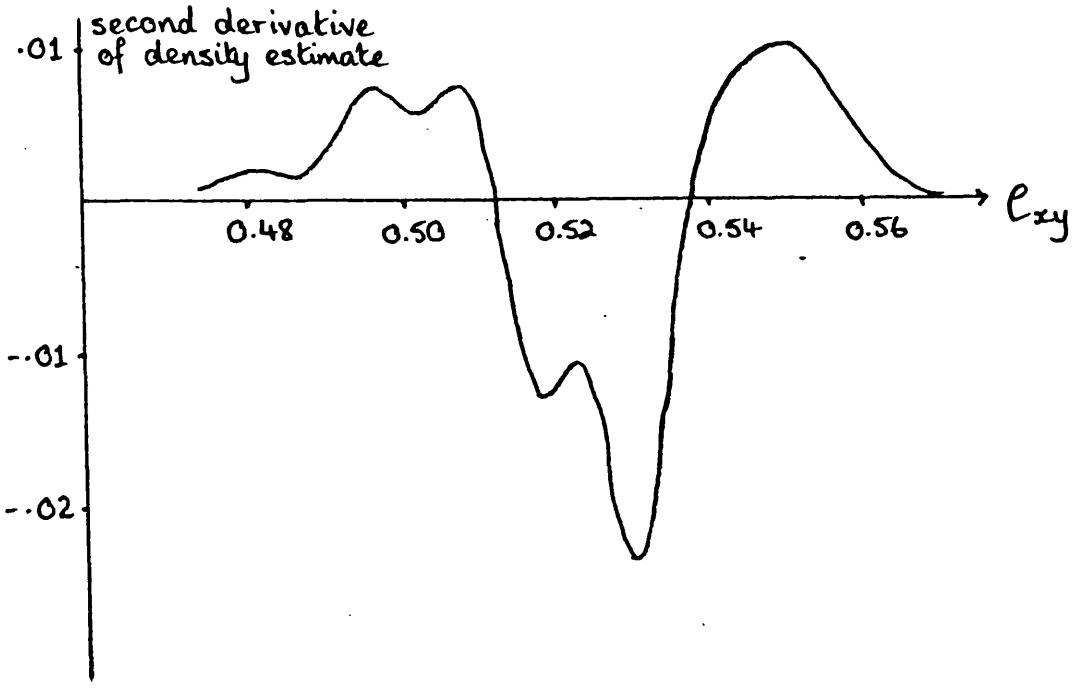
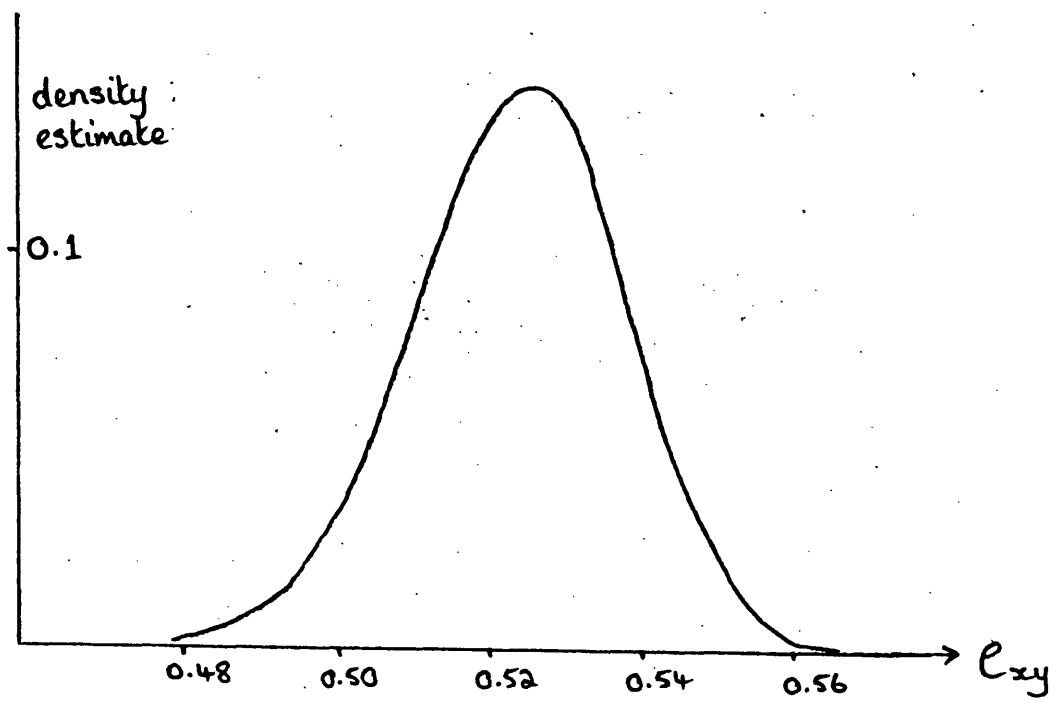


Figure 1b Test graph and density estimate at window width 1.2 for the density function of $\rho_{11.0 \times 11.392}$



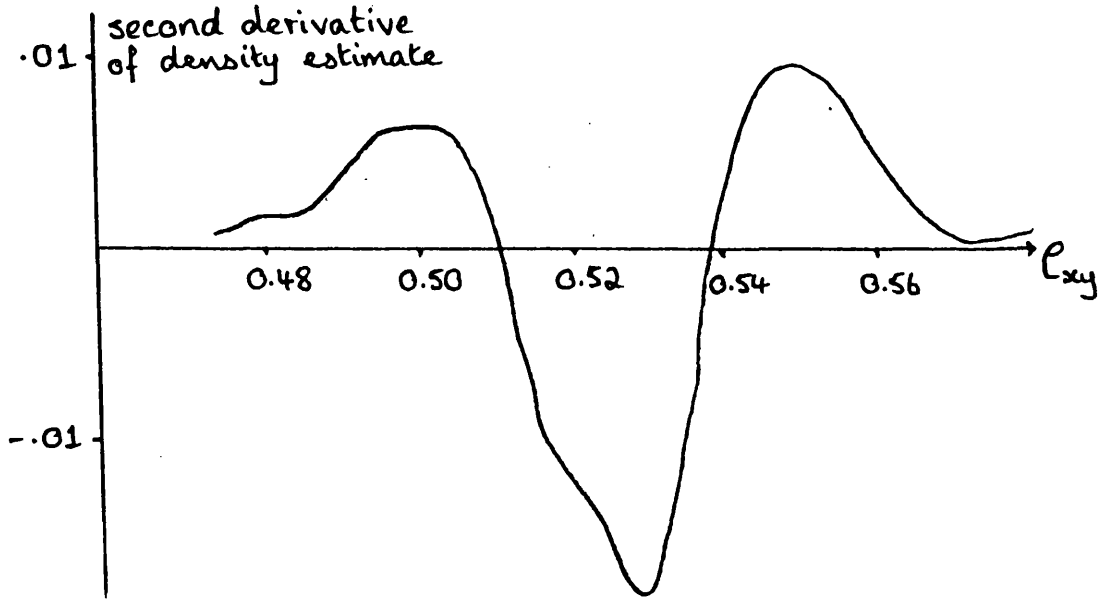
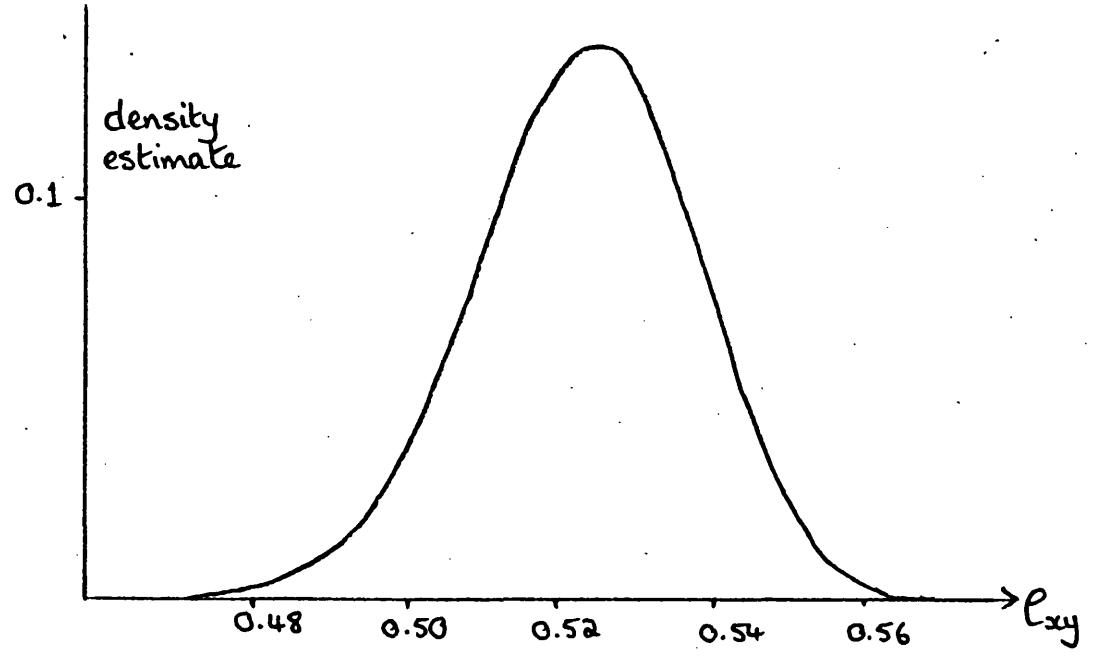


Figure 1c Test graph and density estimate at window width 1.4 for the density function of $\rho_{11.0 \times 11.392}$



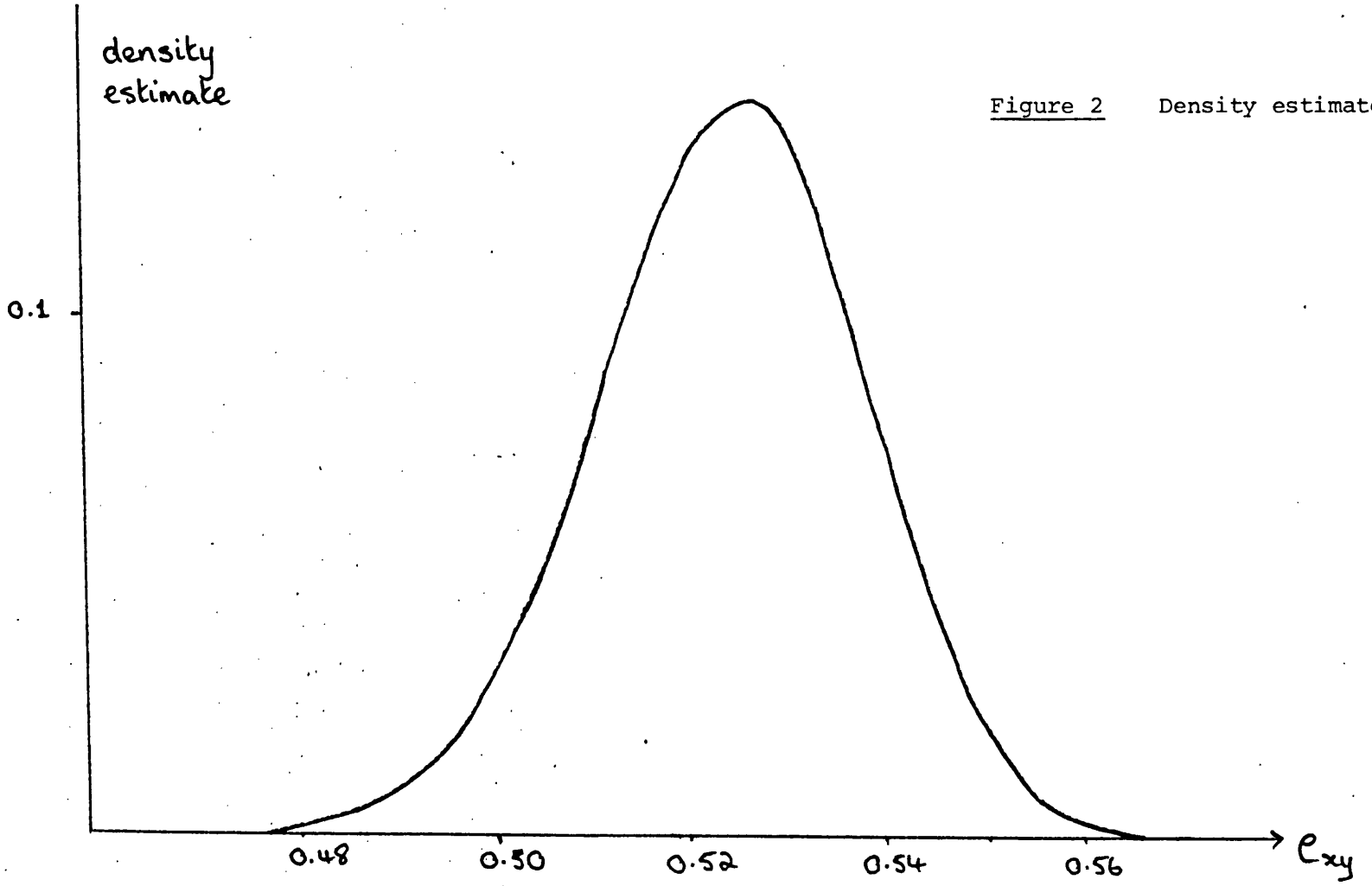


Figure 2 Density estimate of $\rho_{11.0 \times 11.392}$

Figure 3 Density estimate of $\rho_{16.0} \times 16.589$

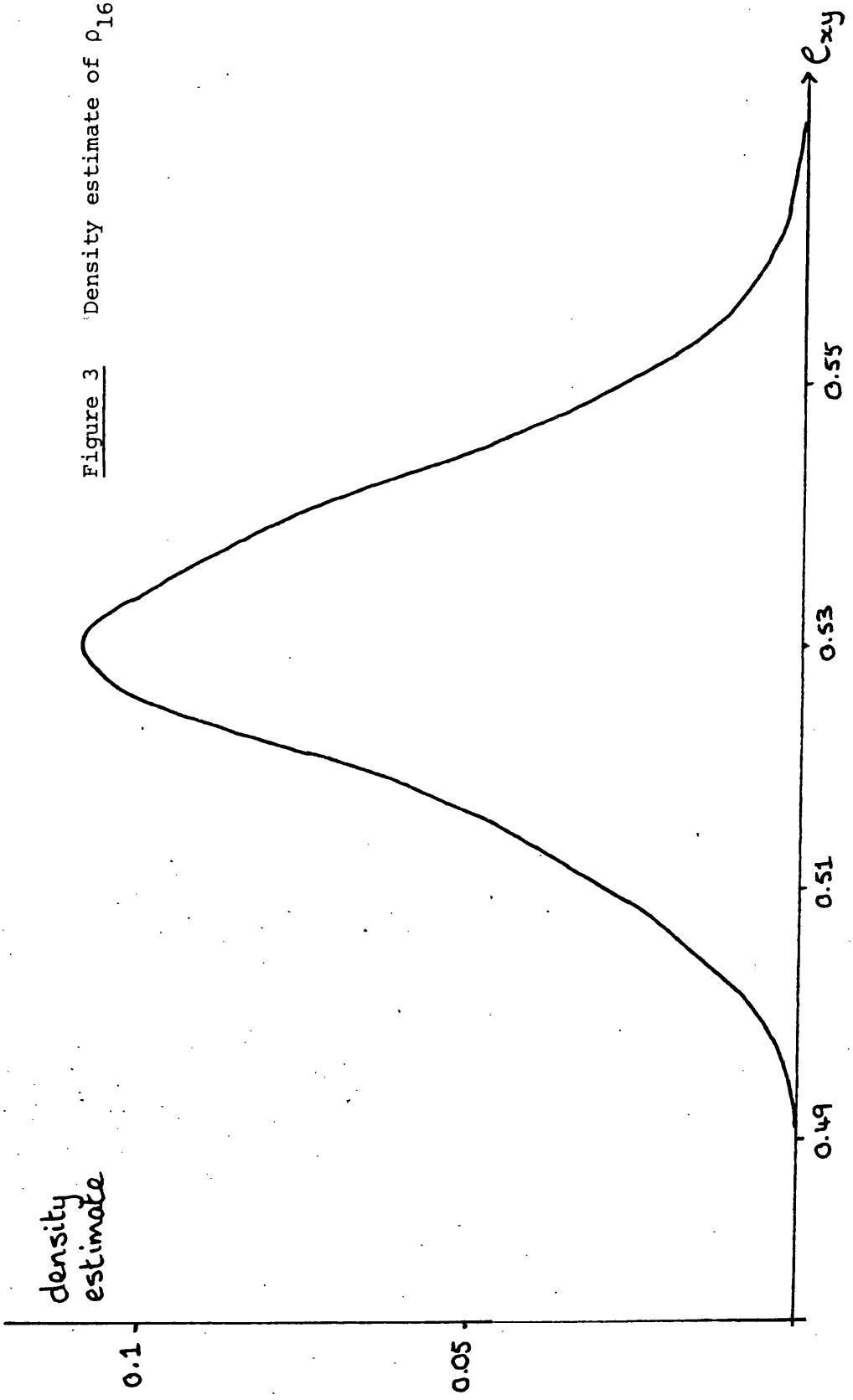
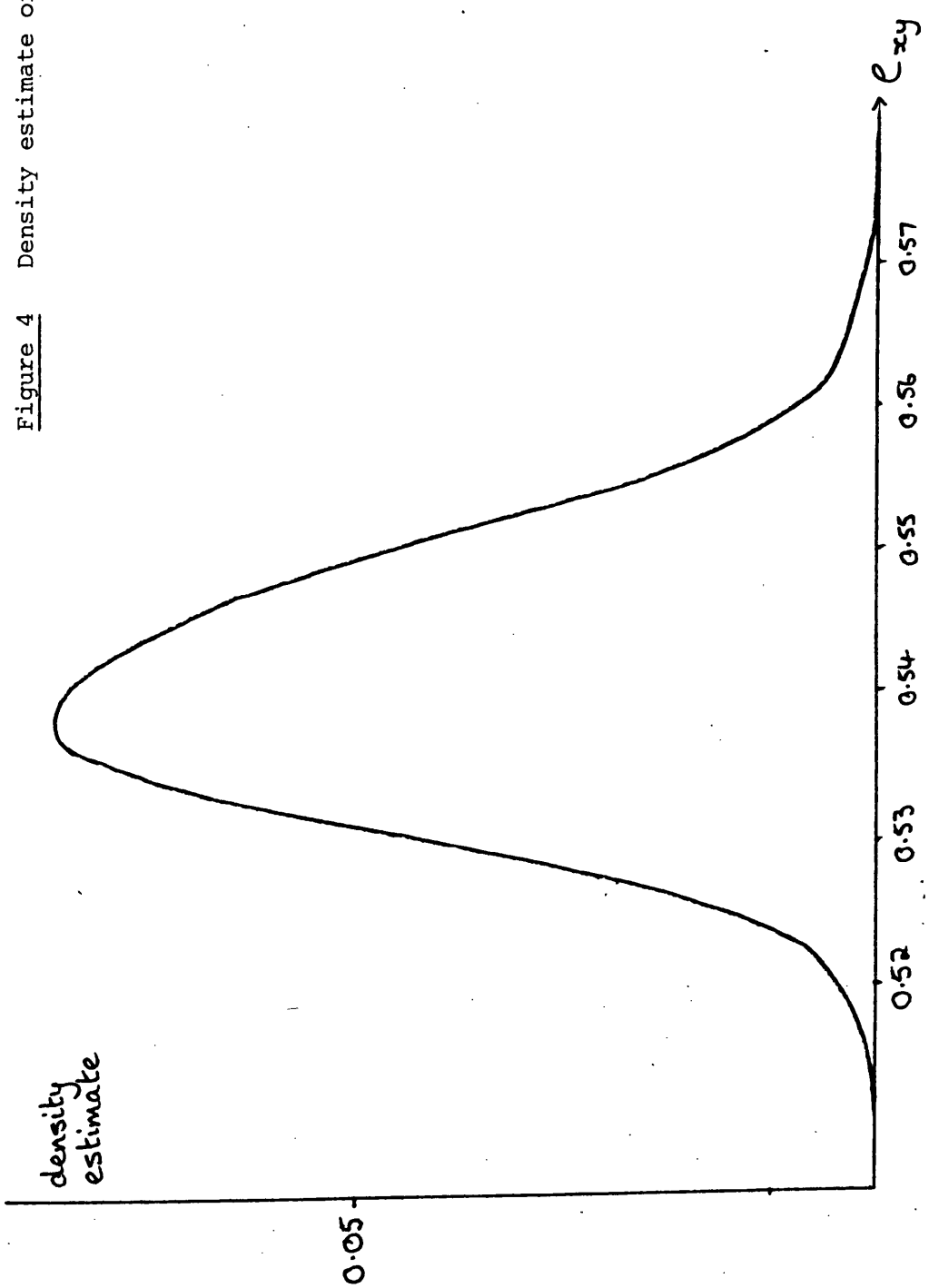


Figure 4 Density estimate of $\rho_{21} \times 21.785$



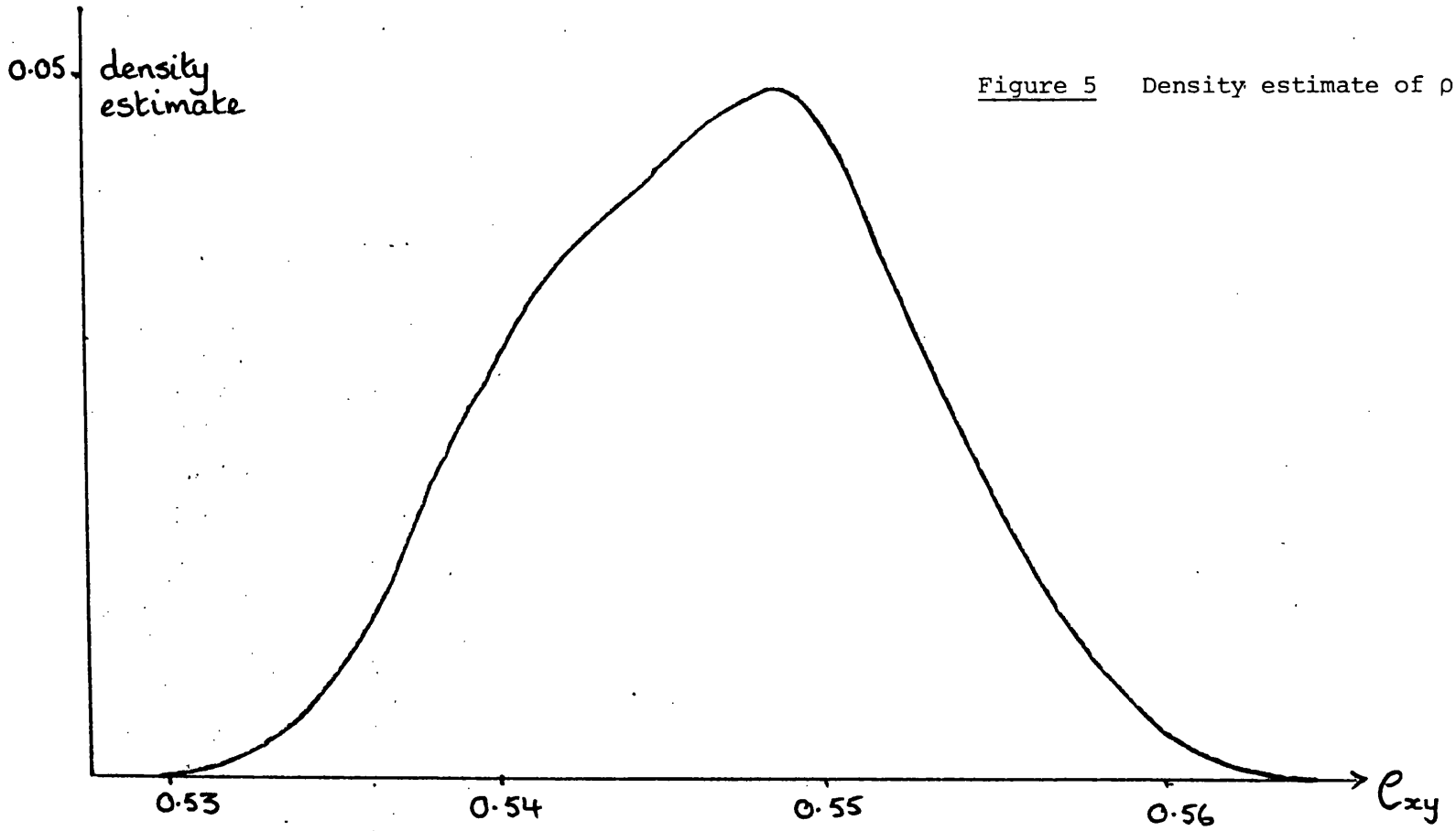


Figure 5 Density estimate of $\rho_{31.0 \times 32.177}$

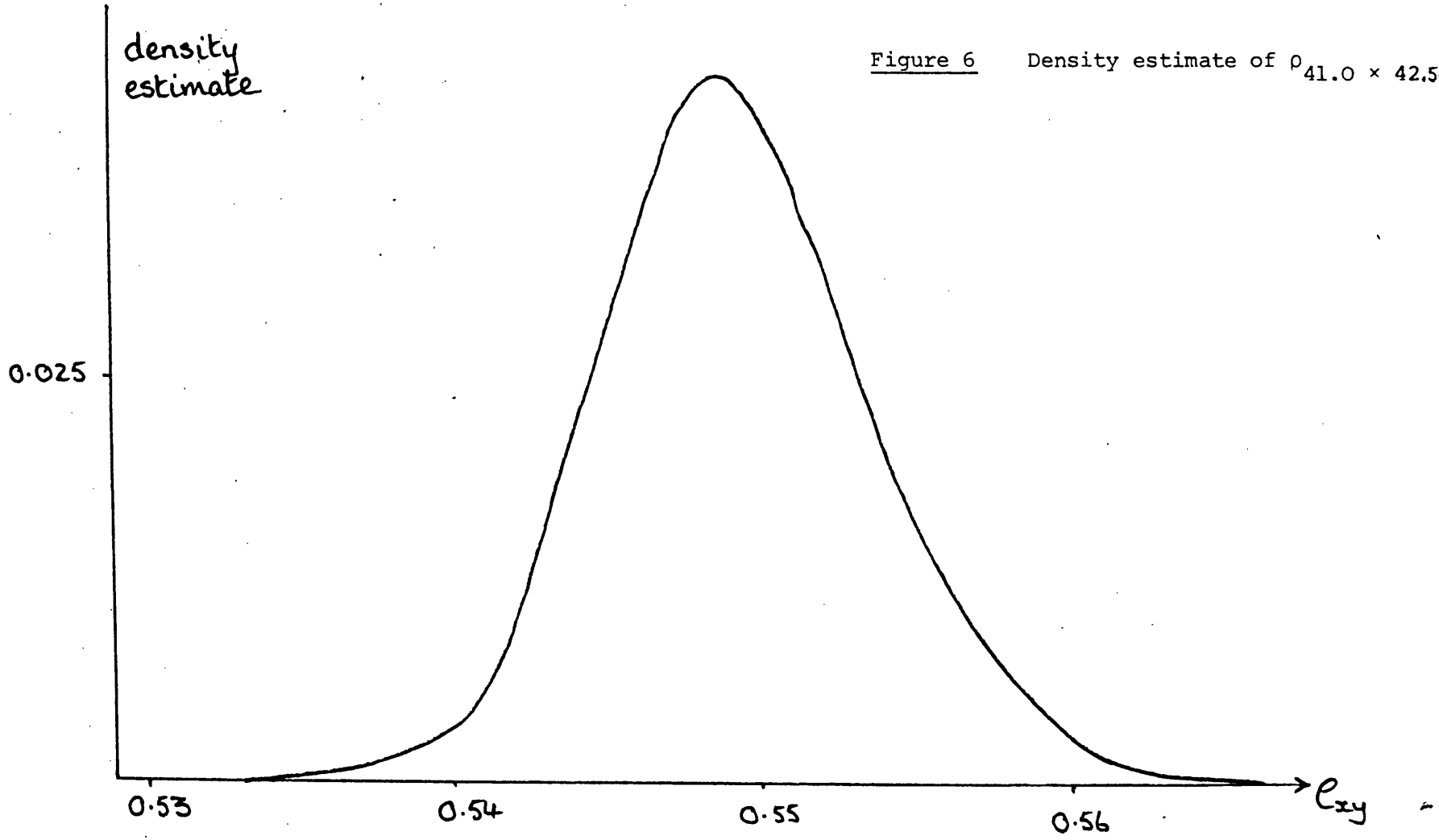


Figure 6 Density estimate of $\rho_{41.0 \times 42.569}$

Figure 7 Density estimate of $\rho_{51} \times 51.23$

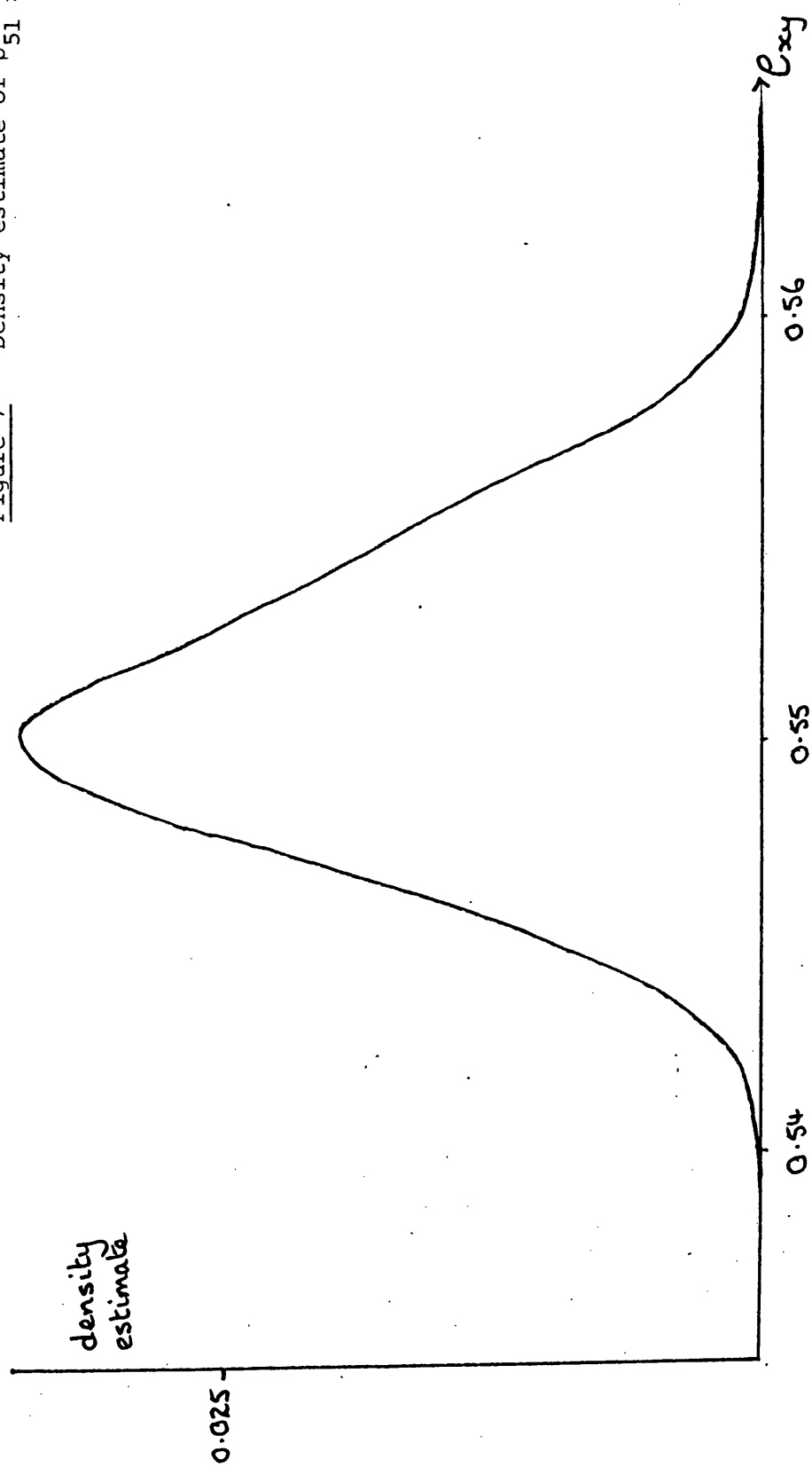
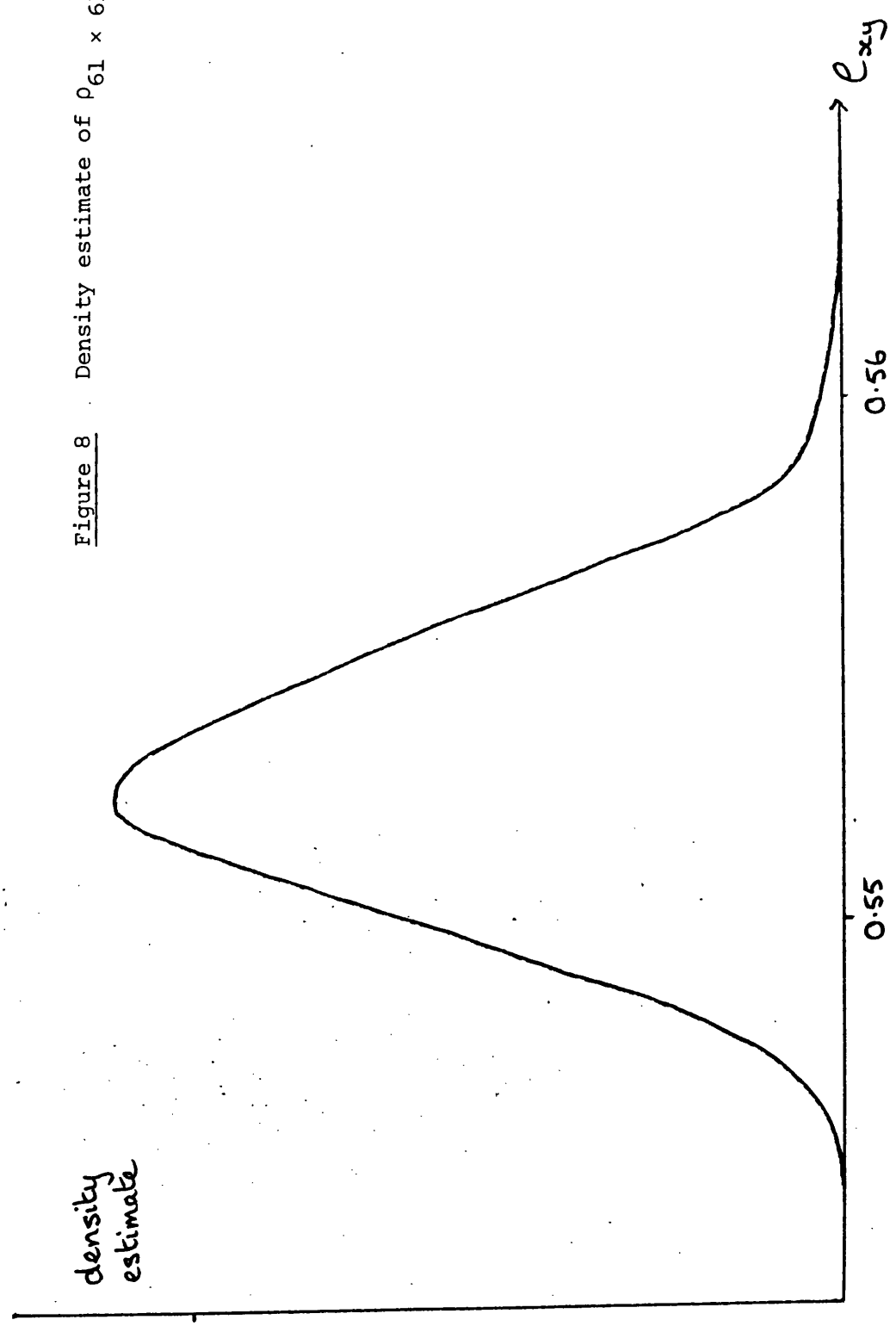


Figure 8 Density estimate of $\rho_{61} \times 61.622$



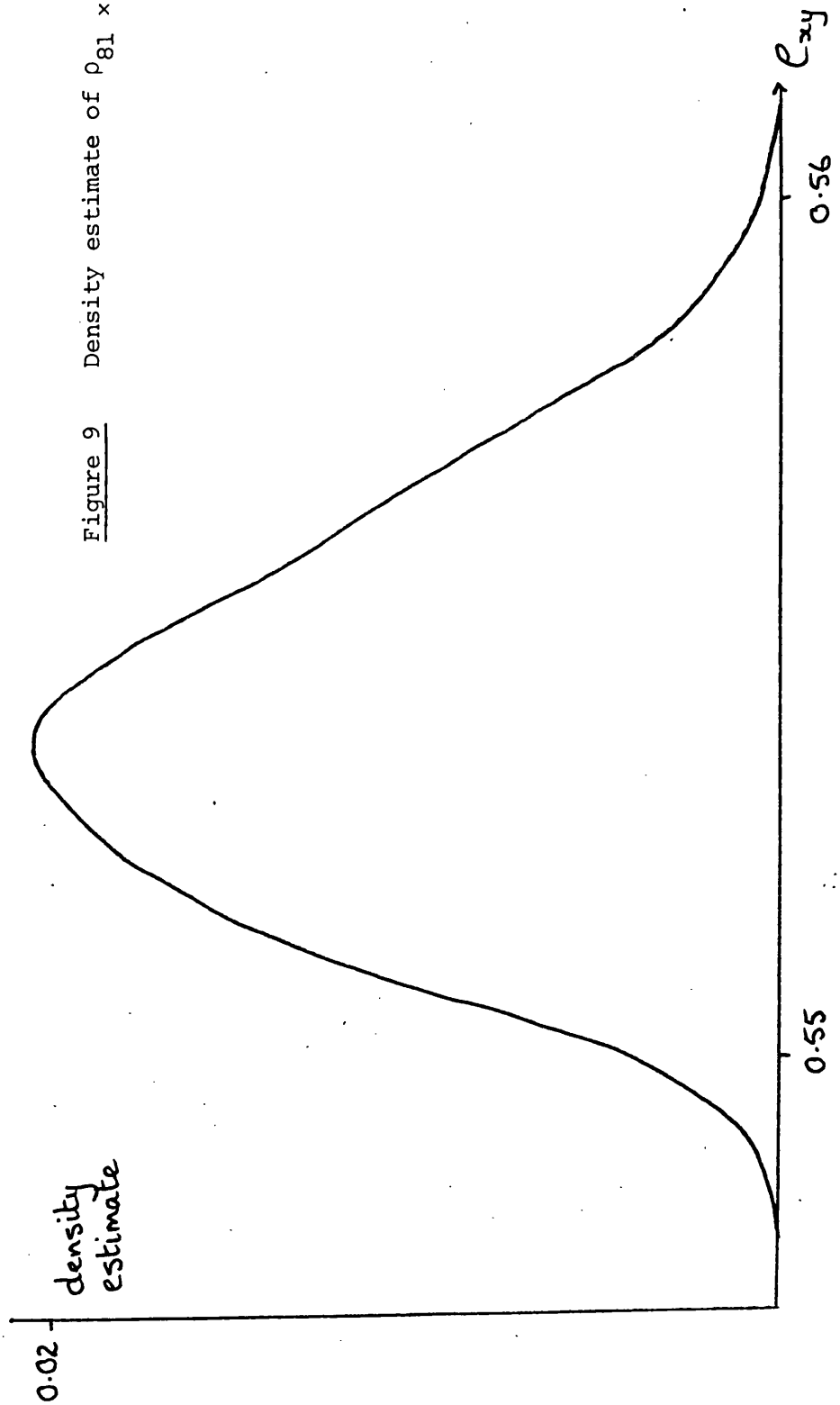


Figure 9 Density estimate of $\rho_{81} \times 82.406$

The eight probability density estimates obtained are shown in Figures 2 to 9. The window widths chosen correspond quite closely to the optimal window widths for a Normal distribution, calculated from (4.14) on page 44 of Wertz (1978). Each of the eight curves obtained, except possibly for the 31.0 × 32.177 rectangle, is consistent with a Normal distribution. Applying goodness-of-fit tests for normality, using the moments of the samples, as described in section 17 of Pearson and Hartley (1966), also confirms that the samples are consistent with having been taken from Normal distributions; the one exception is for the 31.0 × 32.177 rectangle, where the sample indicates kurtosis. It remains an open question as to whether there exists a limiting distribution such that

$$\frac{\rho_{xy} - E\rho_{xy}}{\sigma(\rho_{xy})} \xrightarrow{\mathscr{D}} \rho_{\infty} \quad \text{as} \quad x,y \to \infty$$

Intuitively, one might expect ρ_{∞} to exist and have a standard Normal distribution; this conjecture is supported to some extent by the density estimates in this section, but no results have been proved as yet. One difficulty in attempting to prove such a result comes in establishing a mixing property; even when the sampling region is large, it is not obvious that the SSI process on part of the region is nearly independent of a distant part of the process. This is similar to the difficulty that is encountered in attempting to prove uniqueness of the Mürmann hard core process (see Chapter 4, section 3).

Tanemura (1979) estimated the limiting packing density $\rho = \lim_{x,y\to\infty} E\rho_{xy}$ by assuming that ρ_{xy} has a normal distribution (an assumption supported by the results of this section), by writing

$$\rho_{xy} = \rho + au^{-1} + o(u^{-1})$$

where $u^{-1} = x^{-1} + y^{-1}$, and by assuming that the standard deviation of ρ_{xy} is proportional to u^{-1}. A weighted regression analysis on his ten data points yielded the estimate $\rho = 0.5473 \pm 0.0009$; a similar regression analysis on our eight data points gives an estimate of ρ as 0.5591 ± 0.0010 (95% interval). The two regression lines are shown in Figure 10. The difference between the two estimates is probably due to the fact that Tanemura's algorithm is only approximate; the new algorithm should give a better estimate.

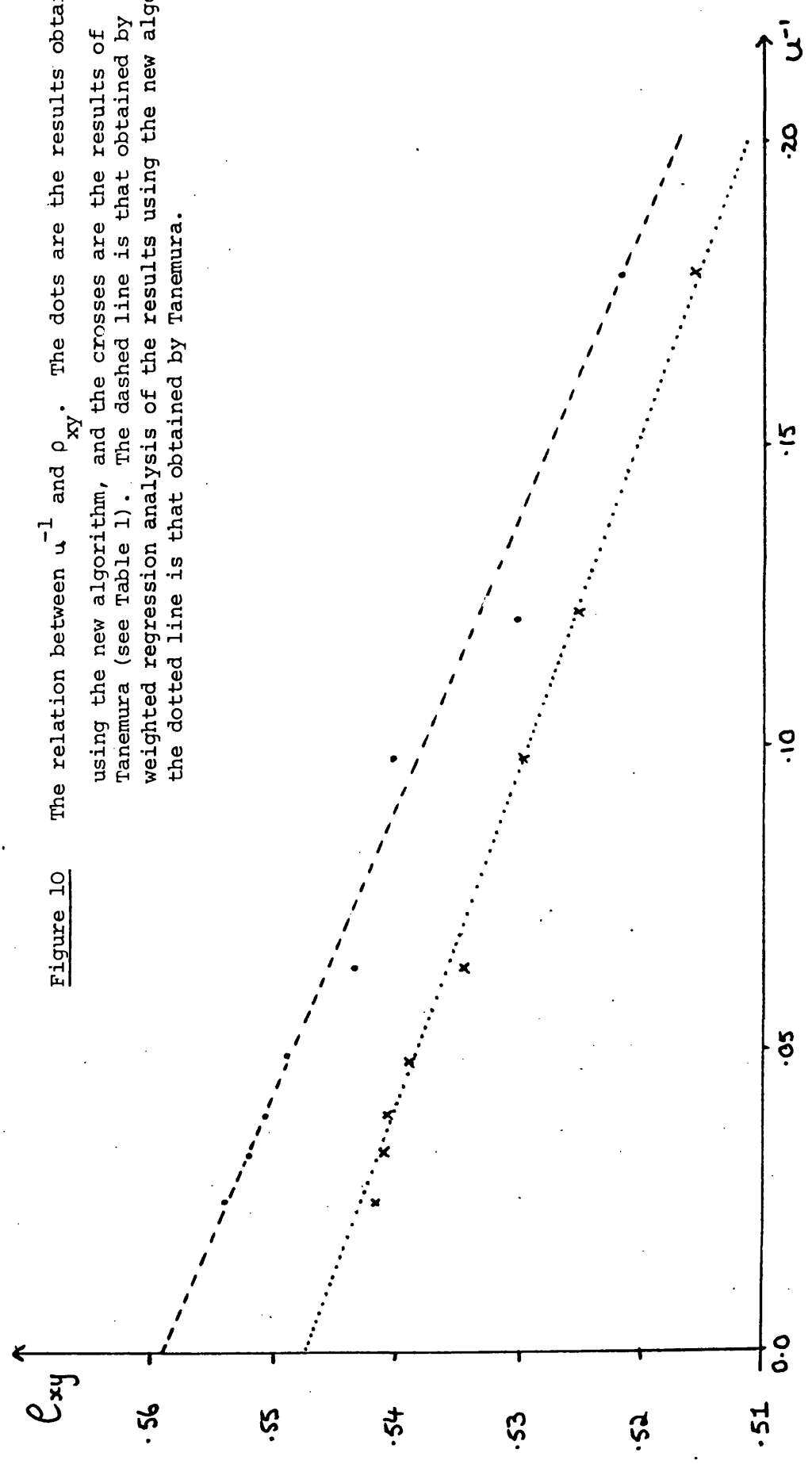


Figure 10 The relation between u^{-1} and ρ_{xy}. The dots are the results obtained using the new algorithm, and the crosses are the results of Tanemura (see Table 1). The dashed line is that obtained by weighted regression analysis of the results using the new algorithm; the dotted line is that obtained by Tanemura.

REFERENCES

AKEDA, Y. and HORI, M. (1975) Numerical test of Palásti's conjecture on two-dimensional random packing density. Nature, 254, 318-319.

AKEDA, Y. and HORI, M. (1976) On random sequential packing in two and three dimensions. Biometrika, 63, 361-366.

ALDER, B.J. and WAINWRIGHT, T.E. (1962) Phase transition in elastic disks. Physical Review, 127, 359-361.

BADDELEY, A. (1980) A limit theorem for statistics of spatial data. Adv. Appl. Probab., 12, 447-461.

BESAG, J.E. (1977) Some methods of statistical analysis for spatial data. Bull. Int. Statist. Inst., 47.

BESAG, J.E. and DIGGLE, P.J. (1977) Simple Monte Carlo tests for spatial pattern. Applied Statistics, 26, 327-333.

BILLINGSLEY, P. (1971) Convergence of Probability Measures. John Wiley and Sons, New York.

BLAISDELL, B.E. and SOLOMON, H. (1970) On random sequential packing in the plane and a conjecture of Palásti. J. Appl. Prob., 7, 667-689.

BROWN, R.T., GEREAU, R.E. and JACOBSEN, R.L. (1981) Measuring the association between pairs of sessile species in a planar continuum, using map data.

BROWN, T.C., MILNE, R.K. and SILVERMAN, B.W. (1981) A class of two-type point processes. Z. Wahrsch. Verw. Gebiete, to appear.

BULINSKII, A.V. and ZHURBENKO, I.G. (1976) A Central Limit Theorem for Additive Random Functions. Theory of Probability and its Applications, 11, 687-697.

CHATFIELD, C. (1975) Statistics for Technology. Chapman and Hall, London.

CHUNG, K.L. (1974) A Course in Probability Theory. Academic Press, London.

COX, D.R. (1965) On the estimation of the intensity function of a stationary point process. J. Roy. Statist. Soc., Ser. B, 27, 332-337.

COX, D.R. (1977) The Role of Significance Tests. Scand. J. Statist, 4, 49-70.

COX, T.F. and LEWIS, T. (1976) A conditioned distance ratio method for analysing spatial patterns. Biometrika, 63, 483-491.

CRANE, M.A. and IGLEHART, D.L. (1974) Simulating Stable Stochastic Systems, I: General Multiserver Queues. J. Assoc. Comp. Mach. 21, 101-113.

DIGGLE, P.J. (1977) The detection of heterogeneity in plant populations. Biometrics, 33, 390-394.

DIGGLE, P.J. (1979) On Parameter Estimation and Goodness-of-fit Testing for Spatial Point Patterns. Biometrics, 35, 87-101.

DIGGLE, P.J. (1980) Statistical methods for spatial point patterns in ecology. In Spatial and Temporal Analysis in Ecology (R.M. Cormack and J.K. Ord, eds.), International Co-operative Publishing House, Maryland.

DIGGLE, P.J., BESAG, J. and GLEAVES, J.T. (1976) Statistical analysis of spatial point patterns by means of distance methods. Biometrics, 32, 659-667.

DIGGLE, P.J. and MILNE, R.K. (1981) Bivariate Cox Processes: some models for bivariate spatial point patterns. Unpublished.

DOOB, J.L. (1953) Stochastic Processes. John Wiley and Sons, New York.

FRYER, M.J. (1977) A review of some non-parametric methods of density estimation. J. Inst. Maths. Applics., 20, 335-354.

GERRARD, D. (1969) Competition quotient: a new measure of the competition affecting individual forest trees. Michigan State University Agricultural Experimental Station, Research Bulletin No. 20.

GLASS, L and TOBLER, W.R. (1971) Uniform distribution of objects in a homogeneous field, cities on a plain. Nature, 233, 67-68.

GREEN, P.J. and SIBSON, R. (1978) Computing Dirichlet Tessellations in the plane. Computer Journal, 21, 168-173.

GUILD, F.J. and SILVERMAN, B.W. (1978) The Microstructure of glass fibre reinforced polyester resin composites. Journal of Microscopy, 114, 131-141.

HALL, P. (1935) On Representations of Subsets. J.L.M.S., 10, 26-30.

HANISCH, K-H., and STOYAN, D. (1979) Formulas for the Second-order Analysis of Marked Point Processes. Mathematische Operationsforschung und Statistik Series Statistics, 10, 555-560.

JODREY, W.S. and TORY, E.M. (1980) Random Sequential Packing in $\mathbb{R}^n$. J. Stat. Comp., 10, 87-93.

KALLENBERG, O. (1975) Random Measures. Academic Press, London.

KELLY, F.P. (1977) Contribution to discussion of Ripley (1977).

KELLY, F.P. and RIPLEY, B.D. (1976) A note on Strauss's model for clustering. Biometrika, 63, 357-360.

KRONMAL, R.A. and PETERSON, A.V. (1979) On the Alias Method for generating random variables from a discrete distribution. Amer. Stat., 33, 214-218.

LEWIS, P.A.W. (1972) Stochastic Point Processes: Statistical Analysis, Theory and Applications. John Wiley and Sons, New York.

LINDVALL, T. (1977) A probabilistic proof of Blackwell's Renewal Theorem. Ann. Probab., 5, 482-485.

MANNION, D. (1964) Random space-filling in one dimension. Public. Math. Inst. Hung. Acad. Sci., 9, 143-153.

MANNION, D. (1976) Random packing of an interval. Adv. Appl. Prob., 8, 477-501.

MATÉRN, B. (1960) Spatial Variation. Meddelanden fran statens skegsforskningsinstitut, Bd. 49, No. 5, Stockholm.

MATTHES, K., KERSTAN, J. and MECKE, J. (1978) Infinitely Divisible Point Processes. John Wiley and Sons, New York.

MOHN, E. and STAVEM, P. (1974) On the Distribution of Randomly Placed Disks. Biometrics, 30, 137-156.

MÜRMANN, M. (1978) Poisson point processes with exclusion. Z. Wahrsch. Verw. Gebiete, 43, 23-37.

PAGE, E.S. (1959) The distribution of vacancies on a line. J.R. Statist. Soc., A, 21, 364-374.

PALÁSTI, I. (1960) On some random space filling problems. Publ. Math. Inst. Hung. Acad. Sci., 5, 353-360.

PALOHEIMO, J.E. (1971) On a theory of search. Biometrika, 58, 61-75.

PEARSON, E.S. and HARTLEY, H.O. (1966) Biometrika Tables for Statisticians, Volume I. Cambridge University Press.

PEEBLES, P.J.E. (1974a) Statistical Analysis of Catalogs of Extragalactic Objects. IV. Cross-correlation of the Abell and Shane-Wirtanen Catalogs. Astrophys. J. Suppl., 28, 37-50.

PEEBLES, P.J.E. (1974b) The nature of the distribution of galaxies. Astron. and Astrophys., 32, 197-202.

PITMAN, J.W. (1974) Uniform rates of convergence for Markov Chain transition probabilities. Z. Wahrsch. Verw. Gebiete, 29, 193-227.

PRESTON, C.J. (1976) Spatial birth-and-death processes. Bull. Inst. Internat. Statist. 46, 371-391.

RAO, C.R. (1952) Advanced Statistical Methods in Biometric Research. John Wiley and Sons, New York.

RÉNYI, A. (1958) On a one-dimensional problem concerning random space filling. Publ. Math. Inst. Hung. Acad. Sci., 3, 109-127.

RIPLEY, B.D. (1976) The second-order analysis of stationary point processes. J. Appl. Prob. , 13, 255-266.

RIPLEY, B.D. (1977) Modelling spatial patterns (with discussion). J. Roy. Statist. Soc., Ser. B., 39, 172-212.

RIPLEY, B.D. (1979a) Simulating spatial patterns: dependent samples from a multivariate density. J. Roy. Statist. Soc., Ser. C., 28, 109-112.

RIPLEY, B.D. (1979b) Tests of "randomness" for spatial point patterns. J. Roy. Statist. Soc., Ser. B., 41, 368-374.

ROGERS, C.A. (1964) Packing and Covering. Cambridge University Press, London.

ROSS, S.M. (1969) A Markovian replacement model with a generalisation to include stocking. Management Science, 15, 702-715.

ROSS, S.M. (1970) Applied Probability Models with Optimization Applications. Holden-Day.

SAUNDERS, R. and FUNK, G.M. (1977) Poisson limits for a clustering model of Strauss. J. Appl. Prob., 14, 776-784.

SILVERMAN, B.W. (1976) Limit theorems for dissociated random variables. Adv. Appl. Prob., 8, 806-819.

SILVERMAN, B.W. (1978a) Choosing the window width when estimating a density. Biometrika, 65, 1-11.

SILVERMAN, B.W. (1978b) Distances on circles, toruses and spheres. J. Appl. Probab., 15, 136-143.

SILVERMAN, B.W. (1982) Kernel density estimation using the Fast Fourier Transform. Applied Statistics, to appear.

SILVERMAN, B.W. and LOTWICK, H.W. (1981) Methods for analysing spatial processes of several types of points. Submitted for publication.

SOLOMON, H. (1967) Random packing density. Proc. Fifth Berkeley Symp. Math. Statist. Prob., 3, 119-134.

STRAUSS, D.J. (1975) A model for clustering. Biometrika, 62, 467-475.

TAKAHASHI, T. (1970) Lobular structure of the human liver from the viewpoint of hepatic vascular architecture. Tohoku Journal of Experimental Medicine, 101, 119-140.

TALBOT, D.R.S. and WILLIS, J.R. (1980) The effective sink strength of a random array of voids in irradiated material. Proc. R. Soc. Lond. A., 370, 351-374.

TANEMURA, M. (1979) On random complete packing by discs. Ann. Inst. Statist. Math., 31, 351-365.

WALKER, A.J. (1977) An efficient method for generating discrete random variables with general distributions. ACM Trans. Math. Soft., 3, 253-256.

WERTZ, W. (1978) Statistical Density Estimation. A Survey. Vandenhoeck and Ruprecht in Göttingen.

WETHERILL, G.B. (1969) Sampling Inspection and Quality Control. Chapman and Hall, London.

WHITE, S.D.M. (1979) The hierarchy of correlation functions and its relation to other measures of galaxy clustering. Mon. Not. R. astr. Soc., 186, 145-154.

WILLIS, J.R. (1980) A polarization approach to the scattering of elastic waves - II. Multiple scattering from inclusions. J. Mech. Phys. Solids., 28, 307-327.

WOODWARD, R.H. and GOLDSMITH, P.L. (1964) Cumulative Sum Techniques. I.C.I. Monograph No. 3., Oliver & Boyd, Edinburgh.

Math. Proc. Camb. Phil. Soc. (1981), **90**, 155
Printed in Great Britain

Convergence of spatial birth-and-death processes

BY H. W. LOTWICK AND B. W. SILVERMAN
University of Bath

(*Received* 19 *September* 1980)

Abstract

Some models for spatial point processes are difficult to simulate directly and are most easily realized as the equilibrium distribution of certain spatial–temporal Markov processes. This paper examines the convergence of such processes, concentrating mainly on the 'hard core' case when the points represent the centres of non-overlapping discs. Coupling methods from the theory of Markov chains are used to establish sufficient conditions for the processes to converge to the required equilibrium, and to give a lower bound on the rate of convergence. One technique used is to couple processes when they become close in a suitable metric.

1. *Introduction*

It is often necessary, for the purposes of testing the goodness of fit of a model, to simulate realizations of various spatial point patterns. Common models for spatial patterns are described by Strauss (**11**, 1975) and Kelly and Ripley (**5**, 1976), where the points lie in a bounded region E, and the joint density is proportional to

$$f(x) = b^n c^s, \tag{1}$$

where n = number of points in the configuration and s = number of pairs of points within distance R of each other.

The case $c = 0$ is one model for a 'hard core' process where the points represent the centres of non-overlapping discs of radius $\frac{1}{2}R$. For simplicity, all point processes considered in this paper will be defined on the unit square U in two-dimensional Euclidean space.

These models are difficult to simulate directly, and are most conveniently simulated using spatial birth and death processes, as described by Ripley (1977 and 1979). In this method, the steps of a Markov process are simulated, the process being chosen so that it has as its equilibrium distribution the required Kelly–Ripley process. By allowing the process to run for a large number of steps, it should get 'near' to equilibrium, and thus the final configuration should be 'close' to what is required.

In this paper, attention is concentrated mainly on the case $c = 0$, i.e. on hard core processes. Coupling arguments from the theory of Markov chains are used to determine conditions under which the processes converge to a unique equilibrium distribution,

0305-0041/81/0000-8450 $03.50

and to give a lower bound on the rate of convergence. Particular use is made of 'ϵ-coupling', which couples two processes when they get within a distance ϵ of each other in a certain metric. This technique bears a certain resemblance to the coupling method introduced by Lindvall (1977). However, there is a fundamental difference: whereas Lindvall's method guarantees only that the coupled processes remain close, our 'ϵ-coupling', if it is successful, ensures that the two processes eventually become identical.

We shall consider two different types of spatial birth and death process. The first of these, the general birth and death process, gives rise to processes of given intensity; the second class considered, the fixed number process, is of interest when conditioning on the total number of points observed.

2. *Description of the processes*

(*a*) *The general birth and death process*

These processes are special cases of spatial Markov birth and death processes as described by Preston(8). Let x denote the state of the process at time t. In an infinitesimal interval $[t, t+\delta t)$, points are added to the process (are 'born') with spatial intensity $B_x(\xi)$ where, for constants $b > 0$ and $c \in [0, 1]$,

$$B_x(\xi) = bc^{t(\xi, x)}; \tag{2}$$

here $t(\xi, x)$ is the number of points of x within R of ξ. For the hard core case, $c = 0$ and so $B_x(\xi) = bI[\xi \text{ is not within } R \text{ of any point of } x]$.

In the interval $[t, t+\delta t)$, points are deleted from the process ('die') independently of one another, each with probability δt. It is easily shown, by theorem 7·1 of Preston(8), that these processes converge to a unique limiting equilibrium process, which is the required process defined in (1) above.

(*b*) *The fixed-number process*

Starting from an initial configuration of N points, single points are alternately deleted and added. At each stage the point to be deleted is chosen uniformly from the existing points, while the point to be added is placed with intensity proportional to $B_x(\xi)$, as defined in (2) above, where x is the process of $N-1$ points remaining after the deletion.

Ripley(10) argues that, for the purposes of computer simulation, the state space is finite, so that, since the Markov chain is aperiodic, convergence of the process is guaranteed provided that the chain is irreducible. It is noted by Ripley(9) that one possible equilibrium distribution is precisely the process described in (1) above conditioned on exactly N points being present. Irreducibility of the chain would guarantee uniqueness of this equilibrium distribution.

We shall consider the convergence of the process without discretizing to a finite state space, give sufficient conditions for irreducibility and investigate rates of convergence to equilibrium.

3. *Convergence of the general birth-and-death process*

Let $X(t)$ be a spatial birth and death process evolving in time with transition rates as given in Section 2*a* and with initial distribution λ. Let $Y(t)$ be a similar chain, started in equilibrium, i.e. having as its initial distribution the equilibrium distribution π of the process. We will investigate the convergence of the process $X(t)$ by considering the double process $Z(t) = (X(t), Y(t))$.

The state space of this double process is almost entirely continuous, but has one 'discrete' state $(0, 0)$, the state where there are no points present in both $X(t)$ and $Y(t)$. Because the state space is partly discrete, the coupling argument of Pitman (7) on discrete state space Markov chains can be modified to apply to the process $Z(t)$. Let τ be the first hit on $(0, 0)$, and define

$$U(t) = \begin{cases} X(t), & t < \tau, \\ Y(t), & t \geqslant \tau. \end{cases}$$

Thus we couple $X(t)$ and $Y(t)$ when they both hit 0. It follows that $U(t) \overset{\mathscr{D}}{=} X(t)$, and standard coupling arguments yield that, for any measurable subset C of the space $\mathscr{B}$ of all possible configurations,

$$|P(X(t) \in C) - \pi(C)| \leqslant P(\tau > t). \tag{3}$$

We thus need to investigate $P(\tau > t)$ to obtain results on the rate of convergence of $X(t)$ to equilibrium. Doing this yields the following

THEOREM A. *In the case $c = 0$, there are constants γ, ρ with $0 < \gamma < \infty$ and $0 < \rho < 1$ such that, for all measurable $C \subseteq \mathscr{B}$,*

$$|P(X(t) \in C) - \pi(C)| \leqslant \gamma \rho^t, \tag{4}$$

i.e. the convergence is geometric.

Proof. Suppose x and y are any configurations. Let $m = |x| + |y|$. Let

$$\beta(x) = \int_U B_x(\xi)\, d\xi \quad \text{and} \quad \beta_1 = \beta(x) + \beta(y).$$

Then, conditional on $Z(0) = (x, y)$, by standard Markov process theory, for some function $K_1(m, t_1)$,

P(next event in Z occurs before time t_1 and is a death)

$= [1 - \exp\{-(\beta_1 + m) t_1\}]\, m/(\beta_1 + m)$

$\geqslant K_1(m, t_1) > 0$ for all m, since $\beta_1 \leqslant 2b$. (5)

It follows that, given $Z(0) = (x, y)$, for $t_0 > 0$,

$P(Z(t)$ hits $(0, 0)$ during $[0, t_0])$

$\geqslant P$(first m events happen before t_0 and are all deaths)

$\geqslant P$(first m inter-event times are at most t_0/m and the first m events are all deaths)

$$\geqslant \prod_{j=1}^{m} K_1(j, t_0/m) = K_0(m, t_0),$$

say, where $K_0(m, t_0) > 0$ by (5) above.

In the hard core case, m is bounded and so there exists $K(t_0) > 0$ such that, for all possible initial configurations,

$$P(Z(t) \text{ hits } (0,0) \text{ during } [0, t_0]) \geqslant K(t_0).$$

By a similar argument to that of Doob (4), p. 194, for any $t > t_0$,

$$P(\tau > t) \leqslant (1-K)^{t/t_0 - 1};$$

putting $\gamma = (1-K)^{-1}$ and $\rho = (1-K)^{1/t_0}$ completes proof of the theorem. |

The inequalities developed in the first part of the proof also apply to the 'soft core' case $c > 0$, but for the moment we shall restrict attention to the hard core case.

The state 0 satisfies the requirements that the process will return to 0 infinitely often and that the mean time between successive returns is finite. Therefore, in principle at least, using the method of Crane and Iglehart (3), the sections of the process between successive returns to zero may be used as independent and identically distributed blocks in the investigation of the equilibrium properties of the process. However, except for small packing densities, application of the Crane–Iglehart method in practice may be infeasible because the mean time between successive returns to zero, while finite, may be extremely large.

We next consider the 'fixed number' process. Since the state $(0,0)$ is no longer accessible, the technique of coupling the two processes X and Y when they are 'close' rather than identical is used. This technique could also be applied to the general process, where it may well give better bounds on γ and ρ than those derived above.

4. *Distances between processes, and ϵ-coupling*

In order to couple processes when they become 'close' it is first of all necessary to define a suitable metric on point patterns. Suppose x and y are two point patterns with the same number of points. Denote by $\mathscr{P}$ the set of (1–1) correspondences between the points of x and those of y. If $x = \{\xi_1, \ldots, \xi_n\}$, define

$$\rho_0(x,y) = \min_{p \in \mathscr{P}} \max_{1 \leqslant i \leqslant n} \{d(\xi_i, p(\xi_i))\},$$

where d denotes Euclidean distance.

For general patterns x and y, let

$$\rho(x,y) = \begin{cases} \min(1, \rho_0(x,y)) & \text{if } x \text{ and } y \text{ have the same number of points,} \\ 1 & \text{otherwise.} \end{cases}$$

It follows easily that ρ is a metric on point patterns. Note that $\rho(x,y)$ is at least as large as the minimum of 1 and the Prohorov distance between the atomic measures generated by x and y. See Billingsley (1) for the definition of Prohorov distance. Once the distance ρ between point patterns has been defined, it is possible to define a distance d_ρ between point processes, viewed as random point patterns; given point processes X and Y, we define

$$d_\rho(X, Y) = \inf\{\eta > 0 : P(\rho(X(\omega), Y(\omega)) > \eta) < \eta\}.$$

Analogously to exercises 8 and 9 of page 94 of Chung (2), if X and Y are close in the metric d_ρ then their 'distributions' will be close in the obvious analogy of the Lévy distance on distributions over the metric space of point patterns with metric ρ.

To define the coupling, given a positive number $\epsilon < 1$ and two birth-and-death processes $\{X_n\}$ and $\{Y_n\}$, we wait until $\rho(X_n, Y_n) < \epsilon$. A (1–1) correspondence can then be set up between the points of the two configurations X_n and Y_n, with points corresponding to each other being a distance at most ϵ apart. The two processes are then coupled as follows:

(i) When deaths occur, corresponding points are deleted.

(ii) The same point is born for each process, provided it lies in the space allowed for both processes; otherwise different points are inserted for the two processes, in which case the coupling fails and the processes subsequently evolve independently.

The coupling will be successful provided no point illegal for either process is inserted before all the points present when the coupling was started are deleted. By making ϵ small, the probability that the coupling is successful can be made as close to 1 as we wish. Details of the coupling are given in full in the next section.

5. *Convergence of the fixed-number process*

Let $\{X_n\}$ and $\{Y_n\}$ be two realizations of the hard core birth-and-death process with N points, as described in section 2 (*b*) above. Each is observed after each birth, so that, for each process, one step consists of a death followed by a birth. Let $\{X_n\}$ have arbitrary initial distribution λ, with $\{Y_n\}$ having initial distribution π_N, the equilibrium distribution of the process. Thus the chain $\{Y_n\}$ is in equilibrium. Couple X_n and Y_n as described in Section 4 above.

Formally, let

$$T_\epsilon = \min\{n: \rho(X_n, Y_n) < \epsilon\}.$$

Define a new process $Z_n^{(\epsilon)} = (V_n, W_n)$, where

(i) $V_n = X_n$ and $W_n = Y_n$ for $n \leqslant T_\epsilon$.

(ii) When $n = T_\epsilon$, let σ be the (1–1) correspondence for which the minimum in the definition of $\rho(X_n, Y_n)$ is attained. The processes then evolve as follows. First kill a point of each process by setting $(x, y) = (V_n \backslash \xi, W_n \backslash \sigma(\xi))$ where ξ is chosen uniformly from V_n. The 'birth' step is a little more complicated.

For any configuration z, define $A_z = \{\zeta \in U: d(\zeta, \eta) > R, \forall \eta \in z\}$.

Let $A_1 = A_x \cap A_y$, $A_2 = A_x \backslash A_y$, $A_3 = A_y \backslash A_x$.

Let m denote Lebesgue measure. Suppose without loss of generality that

$$m(A_2) \geqslant m(A_3).$$

With probability $m(A_1)/m(A_x)$ choose ζ uniformly in A_1, define $\sigma(\zeta) = \zeta$, set

$$(V_{n+1}, W_{n+1}) = (x \cup \zeta, y \cup \zeta)$$

and continue the attempt to couple the processes. Otherwise abandon the attempt to couple the processes and in subsequent steps let the processes evolve independently. To complete this step set $(V_{n+1}, W_{n+1}) = (x \cup \zeta_1, y \cup \zeta_2)$ where ζ_1 is chosen uniformly in A_2 and ζ_2 has p.d.f. proportional to

$$I[A_y] - m(A_y)\, m(A_x)^{-1} I[A_1].$$

It is easily verified that V_n and W_n will individually have the same probability structure as X_n and Y_n respectively. Note that if all the points present at T_ϵ have died before the coupling fails, then the processes will remain identical; in this event call the coupling 'successful'.

It can easily be shown that

$$P(\text{coupling successful}) \geqslant \left(\frac{p}{p+p_\epsilon}\right)^N,$$

where p is a constant and $p_\epsilon = O(\epsilon)$ as $\epsilon \to 0$, so that

$$P(\text{coupling successful}) \to 1 \quad \text{as} \quad \epsilon \to 0. \tag{6}$$

We use ϵ-coupling to prove the following result:

THEOREM B. *Define X_n and Y_n as above, and define d_ρ as in section 4. If $N < 2(1+R/2)(1+\sqrt{3R})/(3\sqrt{3R^2})$, and if $\eta > 0$ then, on a suitable probability space, there exist versions V_n and W_n of X_n and Y_n respectively such that*

$$d_\rho(V_n, W_n) = o(n^{-1/2N+\eta})$$

as n tends to infinity.

It will then follow, as noted in Section 4 above, that the 'ρ-Levy' distance between X_n and Y_n is $o(n^{-1/2N+\eta})$ as n tends to infinity, for all $\eta > 0$.

Before proving the theorem, we need to prove three lemmas. Let

$$D_n = (X_n, Y_n) \quad \text{and} \quad A_\epsilon = \{(x,y): \rho(x,y) < \epsilon\}.$$

LEMMA 1. *Suppose there exist n_0 and $K_\epsilon > 0$ such that, for all possible N-point configurations x and y,*

$$P(D_n \text{ enters } A_\epsilon \text{ during the first } n_0 \text{ steps} \,|\, D_0 = (x,y)) \geqslant K_\epsilon.$$

Then there exist constants α and δ with $0 < \delta < 1$ and $0 < \alpha < \infty$ such that

$$P(T_\epsilon > n) \leqslant \alpha\delta^n, \tag{7}$$

i.e. T_ϵ has a geometric tail.

Proof. Note that $T_\epsilon = \min\{n: D_n \in A_\epsilon\}$. Using an argument similar to that of Doob (4), page 194, we have that

$$P(T_\epsilon > n) \leqslant (1-K_\epsilon)^{n/n_0-1}. \tag{8}$$

Setting $\alpha = (1-K_\epsilon)^{-1}$ and $\delta = (1-K_\epsilon)^{1/n_0}$ proves the lemma. |

For $C \subseteq \mathscr{B}$ and $\eta > 0$, define

$$C^\eta = \{x \in \mathscr{B}: \exists y \in C \text{ with } \rho(x,y) < \eta\}.$$

LEMMA 2. *Under the conditions of Lemma 1, the process X_n converges to a unique equilibrium distribution π_N in the sense that, given $\eta > 0$, there is an n_1 such that, for all measurable $C \subseteq \mathscr{B}$,*

$$\left.\begin{aligned} P(X_n \in C) &\leqslant \pi_N(C^\eta) + \eta \\ \text{and} \qquad \pi_N(C) &\leqslant P(X_n \in C^\eta) + \eta \end{aligned}\right\} \quad \text{for} \quad n \geqslant n_1.$$

Proof. By (6), we may choose $\epsilon < \eta$ such that $P(\epsilon\text{-coupling is unsuccessful}) < \eta/2$. By (7), we can choose an n_1 such that

$$P(T_\epsilon > n) \leqslant \eta/2 \quad \text{for all} \quad n \geqslant n_1.$$

The construction of $V_n^{(\epsilon)}$ and $W_n^{(\epsilon)}$ is such that, for all measurable $C \subseteq \mathscr{B}$,

$$P(X_n \in C) \leqslant P(Y_n \in C^\epsilon) + p_{n,\epsilon}$$

and

$$P(Y_n \in C) \leqslant P(X_n \in C^\epsilon) + p_{n,\epsilon}$$

where $p_{n,\epsilon} = 1 - P(T_\epsilon \leqslant n \text{ and coupling successful})$. Since $P(Y_n \in C) = \pi_N(C)$, and

$$p_{n,\epsilon} \leqslant P(T_\epsilon > n) + P(\epsilon\text{-coupling unsuccessful}),$$

the result now follows from the above and the fact that $C^\epsilon \subseteq C^\eta$.

We can now prove the convergence of the process. To do this, we need to impose a condition on the packing density of the configuration.

LEMMA 3. *Suppose that*

$$N < 2(1+R/2)(1+\sqrt{3}R)/(3\sqrt{3}R^2). \tag{9}$$

Then the process X_n converges to a unique equilibrium distribution π_N in the sense of Lemma 2.

Note. The condition on N ensures that, given any hard core (radius $\frac{1}{2}R$) configuration of N points, we can always add a point somewhere and still satisfy the hard core condition.

Proof of Lemma 3. To prove the lemma, we demonstrate the existence of K_ϵ and n_0 as in Lemma 1. The result then follows from Lemma 2. Let

$$A(M,R) = \inf_x \{m(A_x): |x| = M \text{ and } x \text{ is hard core } (R)\}$$

and

$$B(M,R) = \sup_x \{m(A_x): |x| = M \text{ and } x \text{ is hard core } (R)\}.$$

The condition on N ensures that $\exists \epsilon > 0$ such that $A(N, R+2\epsilon) > 0$. We assume that ϵ has been chosen so that this condition holds, and $A(N-1, R+2\epsilon) > 4\epsilon(1-\epsilon)$. We also assume that $\epsilon < 0{\cdot}3R$. For any configuration z, let $d_{\min}(z) = \min\{d(\xi,\eta): \xi,\eta \in z\}$.

We establish the existence of K_ϵ and n_0 by showing that, for any two N-point configurations x and y, transition is possible from (x,y) into A_ϵ in $6N-15$ steps. This is done by allowing x to evolve into any configuration z, provided

$$d_{\min}(z) > R+2\epsilon \quad \text{and } z \text{ has no points within } \epsilon \text{ of the edge of } U, \tag{10}$$

and by ensuring that y evolves into a configuration that is within ϵ of z. We have

$$P(D_n \text{ hits } A_\epsilon \text{ in the first } 6N-15 \text{ steps} | D_0 = (x,y))$$

$$\geqslant P^{(6N-15)}(x \to \text{some } z \text{ satisfying condition 10})\, P^{(6N-15)}(y \to N_\epsilon^z | x \to z), \tag{11}$$

where $N_\epsilon^z = \{u: \rho(u,z) < \epsilon\}$.

We consider the two expressions on the right-hand side of (11) separately, looking firstly at the second term. Consider any two configurations y and z, with z satisfying condition (10). For any point $\xi \in z$, and $\epsilon < 0{\cdot}3R$, there are at most six y-discs which overlap the circle of radius $R+\epsilon$ centred at ξ. To move from y to N_z^ϵ in $6N-15$ or fewer steps, remove each of these discs in turn, putting it down elsewhere, and put the final disc down so that its centre lies in the circle of radius ϵ centred at ξ. Repeat the process for each of the points of z. For the last point η of z, at most one disc needs to be dealt with; for the penultimate point, at most two discs, etc. Thus, in at most $6N-15$ steps, Y_n can move by this procedure from y to a configuration which lies in N_z^ϵ.

There are $N!$ orders in which we can deal with the points of z, and, for each of the first $N-6$ of these points, there are 6! orders in which to remove the discs near that point. It follows that

$$P^{(6N-15)}(y \to N_z^\epsilon) \geqslant \left\{\frac{\pi\epsilon^2}{B(N-1,R)}\right\}^N N! \left(\frac{6!}{N^6}\left\{\frac{A(N,R+\epsilon)}{B(N-1,R)}\right\}^5\right)^{N-6} \prod_{r=1}^{5} \frac{r!}{N^r}\left\{\frac{A(N,R+\epsilon)}{B(N-1,R)}\right\}^{r-1}$$

$$= C(N,R,\epsilon), \quad \text{say, where} \quad C(N,R,\epsilon) > 0.$$

We now consider the first term on the right-hand side of (11). We wish to move in $6N-15$ steps from x to a configuration z satisfying condition (10). We do this by ensuring that all the original x-points die by the $(6N-15)$th step and that the configuration remaining after $6N-15$ steps satisfies condition (10). To do this, we must have at least N points being born at least $R+2\epsilon$ from each existing point and at least ϵ from the edge of U. We can afford to make 'mistakes' (i.e. allow a point to be born too close to an existing point or to the edge) provided each such point dies by the $(6N-15)$th step.

Letting r be the number of mistakes made, some manipulation yields that the number of different orders in which the $6N-15$ operations can be performed is at least

$$\binom{\alpha}{r} - \binom{\alpha-2}{r-2}\{(5N-15)^3 - 3(5N-15)^2 + 2(5N-15)\}/3$$

$$+ \binom{\alpha-4}{r-4}\{(5N-15)^3 - 3(5N-15)^2 + 2(5N-15)\}^2/12 = f(N,r), \quad \text{say,}$$

where $\alpha = \alpha(N) = (5N-15)(5N-16)/2$. Now, the area available to put points down in 'permitted' positions is at least $A(N-1,R+2\epsilon) - 4\epsilon(1-\epsilon)$. Thus the probability that a birth is 'permissible' is always at least

$$\{A(N-1,R+2\epsilon) - 4\epsilon(1-\epsilon)\}/B(N-1,R) = X, \quad \text{say.}$$

Since each 'mistake' is a birth in the excluded region, followed at some stage by the death of that point, it follows that

$$P^{(6N-15)}(x \to \text{some } z \text{ satisfying condition (10)})$$

$$\geqslant \binom{6N-15}{N} \frac{N!}{N^N} \sum_{r=0}^{5N-15} f(N,r)\, X^{6N-15-r} Y^r = P(N,R,\epsilon), \quad \text{say,}$$

where $Y = (1-X)/N$. Now, $P(N, R, \epsilon) > 0$. Setting

$$n_0 = 6N-15 \quad \text{and} \quad K_\epsilon = C(N, R, \epsilon) \times P(N, R, \epsilon)$$

completes the proof of the lemma. |

We can now prove Theorem B.

Proof of Theorem B. Define V_n and W_n as above. From ϵ-coupling, we know that

$$P(\rho(V_n(\omega), W_n(\omega)) > \epsilon) \leqslant P(T_\epsilon > n) + P(\epsilon\text{-coupling unsuccessful}) = \gamma_\epsilon(n) + \beta_\epsilon, \quad \text{say.}$$

Now,

$$\beta_\epsilon \leqslant 1 - \left(\frac{p}{p+p_\epsilon}\right)^N,$$

where $p_\epsilon = O(\epsilon)$, so there exists $k_1 > 0$ such that $\beta_\epsilon \leqslant k_1 \epsilon$.

From the bounds on $P(T_\epsilon > n)$,

$$\gamma_\epsilon(n) \leqslant \alpha \delta_\epsilon^n,$$

where $\delta_\epsilon = (1-K_\epsilon)^{1/(6N-15)}$ and, in terms of ϵ and N, $K_\epsilon \geqslant k_2 \epsilon^{2N}$ for some $k_2 > 0$. Thus

$$\delta_\epsilon^n \leqslant (1-k_2\epsilon^{2N})^{n/(6N-15)}$$

and so there exists $k_3 > 0$ such that

$$\delta_\epsilon^n \leqslant \exp(-k_3 n \epsilon^{2N}).$$

Thus there exists $k_4 > 0$ such that

$$P(\rho(V_n(\omega), W_n(\omega)) > \epsilon) < k_4 \exp(-k_3 n \epsilon^{2N}) + k_1 \epsilon = u(\epsilon), \quad \text{say.}$$

Since $\alpha = (1-K_\epsilon)^{-1}$ and $K_\epsilon = O(\epsilon^{2N})$, k_4 can be chosen to be independent of ϵ.

It follows immediately that

$$P(\rho(V_n(\omega), W_n(\omega)) > \epsilon) \leqslant \inf_{\epsilon^* \leqslant \epsilon} \{u(\epsilon^*)\} = \tilde{u}(\epsilon), \quad \text{say.}$$

By definition, $\tilde{u}(\epsilon)$ is continuous and non-increasing. Let $\epsilon_0(n)$ be the unique value such that $\tilde{u}(\epsilon_0) = \epsilon_0$. Then it follows at once that

$$d_\rho(V_n, W_n) \leqslant \epsilon_0(n).$$

Given $\eta > 0$, we now show that $\epsilon_0 = o(n^{-1/2N+\eta})$ to complete the proof. Given A, set $\epsilon_n = An^{-1/2N+\eta}$ and $\epsilon_n^* = n^{-1/2N+\eta/2}$. Then, for sufficiently large n, $\epsilon_n^* < \epsilon_n$ and so, by the definitions of $\tilde{u}$ and u,

$$\tilde{u}(\epsilon_n) \leqslant u(\epsilon_n^*) = k_4 \exp(-k_3 n^{N\eta}) + k_1 n^{-1/2N+\eta/2} < \epsilon_n$$

for sufficiently large n. It follows that, for sufficiently large n, $\epsilon_0 < \epsilon_n$, by the definition of ϵ_0. Thus $\epsilon_0 = o(n^{-1/2N+\eta})$, completing the proof of the theorem. |

In this section, ϵ-coupling and its use to prove convergence of the process have been described for the hard core case. The technique can be extended to deal with the soft core case and to prove convergence for these processes; some details are given in Section 7.

6. *Expected waiting time to coupling*

Having used ϵ-coupling to prove the convergence of the hard core process, it is of interest to investigate how long we need to wait until the two processes get within ϵ of each other and the coupling starts.

Interest centres on the behaviour of K_ϵ as $N \to \infty$. If we allow N to tend to infinity, we must also have R (and ϵ) decreasing. We let $\epsilon = \mu R$ for some constant μ. If we assume that we can vary R with N in such a way as to allow us to treat X, $A(N-1, R+\epsilon)$ and $B(N-1, R)$ as constant, and satisfy inequality (9), it can be shown that, for sufficiently large N,

$$K_\epsilon(N, R) \geqslant \frac{HG^N\epsilon^{2N}}{N^{5(N-5)}},$$

where H and G are constants. Using the right-hand side of (8) as an approximation to $P(T_\epsilon > n)$ and comparing with the geometric distribution with corresponding tail probabilities $\delta^n(1-\delta)$ gives the following upper bound on the expected waiting time until the two processes X_n and Y_n get within ϵ of each other:

$$\text{expected waiting time} \doteqdot (6N-15)/K_\epsilon.$$

As $N \to \infty$, this varies approximately as

$$\frac{N^{5N-24}}{BG^N\epsilon^{2N}},$$

where B and G are constants.

7. *Extensions*

For the fixed number process, Ripley (10) suggested always deleting the oldest surviving point instead of choosing the point to be deleted uniformly from the existing points. This device may be applied equally to the general process; in both cases the equilibrium distribution will be unaffected and it seems plausible that convergence to equilibrium will be more rapid.

Our methods of proof are easily generalized to deal with this extension. In the case of the general process, the bounds on the time to coupling will be identical. In the fixed number case, an adaptation of the ϵ-coupling argument can be applied. The distance between configurations x and y will be redefined as

$$\rho_1(x, y) = \max_{1 \leqslant i \leqslant n} d(x^{(i)}, y^{(i)})$$

where $x^{(i)}$ denotes the ith oldest point in the x pattern. Putting $p(x^{(i)}) = y^{(i)}$ in the definition of ρ_0, it is immediate that ρ_0 is dominated by ρ_1, and so coupling when ρ_1 is small will always take at least as long as coupling when ρ_0 is small. Thus, if it is indeed the case that the modified process converges more rapidly, it will be necessary to use methods other than ϵ-coupling to prove this.

For the fixed number process, ϵ-coupling is easily extended to the 'soft core' ($c > 0$)

case. Defining T_ϵ as before, two processes are coupled in a way analogous to that of the hard core case. The only difference comes in the birth part of the step. Defining

$$f_x(\xi) = B_x(\xi) \Big/ \int_{u \in E} B_x(u)\, dm(u),$$

we allow the same point to be born for both the x and y patterns with probability $\int_{u \in E} \min(f_x(u), f_y(u))\, dm(u)$, in which case the new point is inserted with density proportional to $\min(f_x(u), f_y(u))$.

Lemmas 1 and 2 do not depend on the process being hard core, and Lemma 3 is considerably simpler to prove in the soft core case, with condition (9) now being unnecessary. Since P(coupling unsuccessful) has the same order of magnitude as for hard core processes, Theorem B follows, with no restriction on N being necessary.

Extension of ϵ-coupling to some other birth and death processes, such as those used to simulate fixed range interaction models (see Ripley (9) and (10)) is also straightforward.

Dr G. K. Eagleson and the referee made helpful suggestions, which are gratefully acknowledged.

REFERENCES

(1) Billingsley, P. *Convergence of probability measures* (John Wiley, New York, 1968).
(2) Chung, K. L. *A course in probability theory* (Academic Press, New York, 1974).
(3) Crane, M. A. and Iglehart, D. L. Simulating stable stochastic systems. I. General multiserver queues. *J. Assoc. Comput. Mach.* **21** (1974), 103–113.
(4) Doob, J. L. *Stochastic processes* (John Wiley, New York, 1953).
(5) Kelly, F. P. and Ripley, B. D. A note on Strauss's model for clustering. *Biometrika* **63** (1976), 357–360.
(6) Lindvall, T. A probabilistic proof of Blackwell's Renewal Theorem. *Ann. Probability* **5** (1977), 482–485.
(7) Pitman, J. W. Uniform rates of convergence for Markov chain transition probabilities. *Z. Wahrsch. Verw. Gebiete* **29** (1974), 193–227.
(8) Preston, C. J. Spatial birth-and-death processes. *Bull. Inst. Internat. Statist.* **46** (1976), 371–391.
(9) Ripley, B. D. Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc., Ser. B* **39** (1977), 172–212.
(10) Ripley, B. D. Simulating spatial patterns: dependent samples from a multivariate density. *J. Roy. Statisti. Soc., Ser. C* **28** (1979), 109–112.
(11) Strauss, D. J. A model for clustering. *Biometrika* **62** (1975), 467–475.